

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 25, 2003, 21:56:10 ; Search time 4670.87 Seconds

(without alignments)
11088.207 Million cell updates/sec

Title: US-09-717-321A-2

Perfect score: 1266

Sequence: 1 ttttttttttttttca.....gagcaagaatagggtttt 1266

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: GenBank:*

2: gb_ba:*

3: gb_hg:*

4: gb_in:*

5: gb_ov:*

6: gb_pat:*

7: gb_ph:*

8: gb_pl:*

9: gb_pr:*

10: gb_ro:*

11: gb_sts:*

12: gb_sy:*

13: gb_un:*

14: gb_vl:*

15: em_ba:*

16: em_fun:*

17: em_hum:*

18: em_in:*

19: em_mu:*

20: em_om:*

21: em_or:*

22: em_ov:*

23: em_pat:*

24: em_ph:*

25: em_pl:*

26: em_ro:*

27: em_sts:*

28: em_un:*

29: em_vl:*

30: em_hg_hum:*

31: em_hg_inv:*

32: em_hg_other:*

33: em_hg_mus:*

34: em_hg_pln:*

35: em_hg_rnd:*

36: em_hg_mam:*

37: em_hg_vrt:*

38: em_sy:*

39: em_hgo_hum:*

40: em_hgo_mus:*

41: em_hgo_other:*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1266	100.0	1266	6	AX163738
2	1227.4	97.0	217700	6	AX106124
3	1017	80.3	1017	6	AX163751
4	878.2	69.4	2319	10	BC003828
5	868.2	68.6	2281	10	BC051053
6	862	68.1	269081	2	AC068493
7	445.6	35.2	2302	9	BC050687
8	441.8	34.9	28567	9	HSN132695
9	441.8	34.9	212827	9	AC009412
10	430.6	34.0	2315	9	AK054993
11	420.2	33.2	1232	6	AX408985
12	420.2	33.2	1232	9	HUMPO2ST9
13	408.2	32.2	185745	2	AC102775
14	408.2	32.2	202342	10	AL583884
15	369.4	29.2	192898	2	AC105979
16	369.4	29.2	230015	2	AC132602
17	348	27.5	348	6	AX163753
18	337.2	26.6	179685	10	AC126556
19	309.6	24.5	137625	9	AC104663
20	306.4	24.2	5544	9	AF542527
21	290.6	23.0	1022	6	AX163752
22	250.4	19.8	87616	2	AC139405
23	194.8	15.4	202565	9	AL354696
24	193.2	15.3	605	6	AX341407
25	187.8	14.8	352	6	AX163754
26	175.6	13.9	240973	2	AC123247
27	175.6	13.9	241048	2	AC129824
28	170.6	13.5	262	11	G31709
29	154	12.2	455	11	G26995
30	131.4	10.4	228121	2	AC133022
31	131.4	10.4	239768	2	AC112582
32	131.4	10.4	245468	2	AC130746
33	123	9.7	123	6	AX163737
34	118	9.3	137625	9	AX104663
35	116.4	9.2	110816	9	AC002404
36	112.4	8.9	64781	2	AC102025
37	110.6	8.7	101584	2	CNS01DS5
38	110.6	8.7	155943	9	AC012364
39	105.2	8.3	3544	9	AF542527
40	98.4	7.8	170839	2	AC133467
41	86.6	6.8	180303	9	AL672045
42	86.6	6.8	201012	2	AC021189
43	85.6	6.8	495	6	AX341256
44	84.6	6.7	174316	2	AC022647
45	84.6	6.7	217249	9	AC009902

ALIGNMENTS

RESULT 1

AX163738

LOCUS AX163738 1266 bp DNA linear PAT 22-JUN-2001

DEFINITION Sequence 2 from Patent WO0138579.

ACCESSION AX163738

VERSION AX163738.1 GI:14544844

KEYWORDS

SOURCE

ORGANISM

Rattus norvegicus (Norway rat)

Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

Rattus.

REFERENCE

1 Gould-Rothberg,B.E., Dippio,V.A., Ramsch,T.M. and Gerwein,R.W.

AUTHORS

TITLE

Method of Identifying toxic agents using naid-induced differential

gene expression in liver
 Patent: WO 0138579-A 2 31-MAY-2001.
 Curagen Corporation (US)
 Location/Qualifiers
 FEATURES
 source
 1. 1266
 /organism="Rattus norvegicus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10116"
 BASE COUNT 385 a 258 c 285 g 338 t
 ORIGIN
 Query Match 100.0%; Score 1266; DB 6; Length 1266;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1266; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 TTTTCTTTTCTTTTCTTTTCAAGTTCACAAAGACATTTTCTTTTCTTTTATGATTCAGG 60
 |||||
 1 TTTTCTTTTCTTTTCTTTTCAAGTTCACAAAGACATTTTCTTTTCTTTTATGATTCAGG 60
 |||||
 61 ATTATATTAAGTCAATATGCAAAACATCTGCTAATGCTATGCAAAAGATCAATGTAA 120
 |||||
 61 ATTATATTAAGTCAATATGCAAAACATCTGCTAATGCTATGCAAAAGATCAATGTAA 120
 |||||
 121 AAACACTCCACATTTCTGCACATCTGCAATTTGAAAGATTTGTTCTAGTGTGAAAGGC 180
 |||||
 121 AAACACTCCACATTTCTGCACATCTGCAATTTGAAAGATTTGTTCTAGTGTGAAAGGC 180
 |||||
 181 CCACACTGTGTTCTGCGAGTGTAGTGTGTATGCAAGAGCGGCTTACAGCTAGCGCTT 240
 |||||
 181 CCACACTGTGTTCTGCGAGTGTAGTGTGTATGCAAGAGCGGCTTACAGCTAGCGCTT 240
 |||||
 241 GACAGAACCTCCACAGACCCCAAGAGTACCGGAGCATGTGTCCGCTGGGTGAGGTCTAGA 300
 |||||
 241 GACAGAACCTCCACAGACCCCAAGAGTACCGGAGCATGTGTCCGCTGGGTGAGGTCTAGA 300
 |||||
 301 GGGGGGGGCAATCAATGACATGACAGTGTGTGCTGCTGCGACAGACAGTGTGTTTCAGAA 360
 |||||
 301 GGGGGGGGCAATCAATGACATGACAGTGTGTGCTGCTGCGACAGACAGTGTGTTTCAGAA 360
 |||||
 361 TATCTATTAATAGTCTTAAAGCTTAAAGCCGACAGCTATTTCTTACACCCAGTTACTA 420
 |||||
 361 TATCTATTAATAGTCTTAAAGCTTAAAGCCGACAGCTATTTCTTACACCCAGTTACTA 420
 |||||
 421 GAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 480
 |||||
 421 GAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 480
 |||||
 481 TACTATCTACCAAAAAATCTCCGAATGCAATTCAGAAAAGATCTTATATGACAGTCTAG 540
 |||||
 481 TACTATCTACCAAAAAATCTCCGAATGCAATTCAGAAAAGATCTTATATGACAGTCTAG 540
 |||||
 541 ACATATTTGCTCTTTAAGAGAGGGGGTCTTAAAGAAAAGCACTTCTTAAGTACAGACTGTG 600
 |||||
 541 ACATATTTGCTCTTTAAGAGAGGGGGTCTTAAAGAAAAGCACTTCTTAAGTACAGACTGTG 600
 |||||
 601 AGGATGGCCAGTTTAAATATGAGCTCAAGCCCATCTGGGAGGAGGAGGAGGAGGAGGAGG 660
 |||||
 601 AGGATGGCCAGTTTAAATATGAGCTCAAGCCCATCTGGGAGGAGGAGGAGGAGGAGGAGG 660
 |||||
 661 GGGGGGCTCAAGAGAGCACTGATAGATGGCCATTTGTCATCTACTGTTTGAAGAAAT 720
 |||||
 661 GGGGGGCTCAAGAGAGCACTGATAGATGGCCATTTGTCATCTACTGTTTGAAGAAAT 720
 |||||
 721 TAAACCTTTAAAGAGCTTTTACCCCTGACACTTTTATTCAGTGAATTAATCCTCATGTAAAT 780
 |||||
 721 TAAACCTTTAAAGAGCTTTTACCCCTGACACTTTTATTCAGTGAATTAATCCTCATGTAAAT 780
 |||||
 781 GTAGTAAATTAATCTCTACTCATATTAATGCAAAATAGTCTGTCTCTTTGATGAC 840
 |||||
 781 GTAGTAAATTAATCTCTACTCATATTAATGCAAAATAGTCTGTCTCTTTGATGAC 840
 |||||
 841 GTCTGTCTTACACACTCCACAGACACCCAGCTAGGAAGAGATCTTCTGTAGA 900
 |||||

Db 841 GTCTGTCTTACACACTCCACAGACACCCAGCTAGGAAGAGATCTTCTGTAGA 900
 |||||
 QY 901 GGCACACAGAGACCCAGAGTCTTCTTCAAGACCTGACAGAGCCGCTGACGTGATTTTA 960
 |||||
 Db 901 GGCACACAGAGACCCAGAGTCTTCTTCAAGACCTGACAGAGCCGCTGACGTGATTTTA 960
 |||||
 QY 961 GAAATCTCACTAATGAATCAAAAGAGAGCTTTACACCATCTGACGTACAGTACAA 1020
 |||||
 Db 961 GAAATCTCACTAATGAATCAAAAGAGAGCTTTACACCATCTGACGTACAGTACAA 1020
 |||||
 QY 1021 AGTTAGTAAATGAGCATGGGCTGATTAAGTATGACAGGTGCTTACATGACAGCGTCAATTA 1080
 |||||
 Db 1021 AGTTAGTAAATGAGCATGGGCTGATTAAGTATGACAGGTGCTTACATGACAGCGTCAATTA 1080
 |||||
 QY 1081 AGGAGCGTGTCTGTGTACACAGCTGTGGAGCTACGAGAGGCTTGCACCCCTGAGGCC 1140
 |||||
 Db 1081 AGGAGCGTGTCTGTGTACACAGCTGTGGAGCTACGAGAGGCTTGCACCCCTGAGGCC 1140
 |||||
 QY 1141 AGAAGCTGCACTTCTTCAAGCAAAAGTCTCAACAGTGTAGTGTCTTACAGTGTCTCA 1200
 |||||
 Db 1141 AGAAGCTGCACTTCTTCAAGCAAAAGTCTCAACAGTGTAGTGTCTTACAGTGTCTCA 1200
 |||||
 QY 1201 GCACAAAGCACTTCTTCAAGCAAAAGTGTGCAATTTCTTAATCTGACCAAGAAATAGG 1260
 |||||
 Db 1201 GCACAAAGCACTTCTTCAAGCAAAAGTGTGCAATTTCTTAATCTGACCAAGAAATAGG 1260
 |||||
 QY 1261 GATTTT 1266
 |||||
 Db 1261 GATTTT 1266
 |||||

RESULT 2
 AC106124/c 217700 bp DNA linear HTG 10-MAY-2003
 LOCUS Rattus norvegicus clone CH230-119e11, *** SEQUENCING IN PROGRESS
 DEFINITION *** 11 unordered pieces.
 AC106124
 AC106124.6 GI:30521557
 ACCESSION HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
 VERSION
 KEYWORDS Rattus norvegicus (Norway rat)
 SOURCE
 ORGANISM
 Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae;
 Rattus.

REFERENCE
 AUTHORS
 1 (bases 1 to 217700)
 Muzny,D.,Marle, Metzker,M.,Lee, Abramson,S., Adams,C., Alder,J.,
 Allen,C., Allen,H., Alsbrooks,S., Amin,A., Angulano,D.,
 Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
 Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
 Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
 Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
 Cardenas,V., Carter,K., Cavazos,I., Cesari,R., Center,A.,
 Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,A.,
 Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
 Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
 Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Diya,K.,
 Draper,H., Dugan-Kochia,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
 Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
 Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
 Fraser,C.M., Gabel,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
 Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,M., Guevara,W.,
 Gunaratne,P., Haaland,W., Hamill,C., Hamilton,C., Hamilton,J.,
 Harvey,Y., Havlik,P., Hayes,A., Henderson,N., Hernandez,J.,
 Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
 Hollins,B., Howells,S., Huily,S., Hume,J., Idlebird,D., Jackson,A.,
 Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolyet,A.,
 Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
 Kowis,C., Kraft,C.L., Lebow,H., Lervan,J., Lewis,L., Li,Z., Liu,J.,
 Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,D.,
 Lorensonuwa,L., Louie,D., Lozada,R.J., Lu,X., Ma,J.,
 Maheshwari,M., Mahindartine,M., Mahmoud,M., Malloy,K., Mangum,A.,
 Mangum,B., Mapua,P., Martin,K., Martin,K., Martineau,E.,
 Mahoney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,

TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT

Mijalovljivic,A., Miner,G., Munja,E., Montemayor,J., Moore,S.,
 Morgan,M., Morris,K., Morris,S., Mundasa,M., Murphy,M., Natr,L.,
 Nankevicius,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Parks,K.,
 Paetzelknecht,O., Okwono,G., Olairpusaagon,A., Pal,S.,
 Pasteknecht,O., Paul,H., Perez,A., Perez,L., Pfankoch,C.,
 Plappier,F., Polndexter,A., Popovic,D., Prims,E., Pu,L.,
 Puzo,M., Quirroz,J., Rachlin,E., Reeves,K., Reiter,M.A., Reigh,R.,
 Relly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
 Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.,
 Sanders,M., Saverly,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
 Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smaj,D.,
 Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
 Steimle,M., Strong,R., Sutton,A., Swalek,A., Taber,P., Taylor,C.,
 Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Uemari,K.,
 Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
 Wang,Q., Wang,S., Warren,J., Warren,R., Wel,X., White,F.,
 Williams,G., Willson,R., Wlecyk,R., Wooden,H., Worley,K.,
 Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
 Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
 Niederhauser,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
 Weinstein,G. and Gibbs,R.A.
 Direct Submission
 Unpublished
 2 (bases 1 to 217700)
 Direct Submission
 Submitted (12-JAN-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 217700)
 Rat Genome Sequencing Consortium.
 Direct Submission
 Submitted (10-MAY-2003) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On May 10, 2003 this sequence version replaced gi:25094619.
 The sequence in this assembly is a combination of BAC based reads
 and whole genome shotgun sequencing reads assembled using Atlas
 (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described
 in the feature table below represents a scaffold in the Atlas
 assembly (a 'contig-scaffold'). Within each contig-scaffold,
 individual sequence contigs are ordered and oriented, and separated
 by sized gaps filled with Ns to the estimated size. The sequence
 may extend beyond the ends of the clone and there may be sequence
 contigs within a contig-scaffold that consist entirely of whole
 genome shotgun sequence reads. Both end sequences and whole genome
 shotgun sequence only contigs will be indicated in the feature
 table.

----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: GHIL
 Center clone name: CH230-119E11
 ----- Summary Statistics
 Assembly program: Atlas 3.0;
 Consensus quality: 185725 bases at least Q40
 Consensus quality: 190140 bases at least Q30
 Consensus quality: 192842 bases at least Q20
 Estimated insert size: 194758; sum-of-contigs estimation
 Quality coverage: 5x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/genbank/draft_data.html)
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 11 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will

Query Match	Best Local Similarity	Matches 1237; Conservative	Score 97.08; 98.7%;	Score 1227.4; 98.7%;	DB 2; 1.6e-112;	Length 217700;	Indels 16;	Gaps 0;
14	TTTTCAGTCCCAAGACATTTT	TTTTTTT	TTTTTATGATCAAGATTTATAGTCA	73				
64694	TTGTCAGTCCCAAGACATTTT	TTTTTTT	TTTTTATGATCAAGATTTATAGTCA	64635				
74	TACATGCAAAACATCTACTG	TACTACTG	TATGACAAAGATCAATGTAACCA	133				
64634	TACATGCAAAACATCTACTG	TACTACTG	TATGACAAAGATCAATGTAACCA	64575				
134	TTTCGCAACTGTCATTTGAAA	AAAGTTGTTCTT	CTAGTGTCGAAAGGCCCACTGTGT	193				
64574	TTTCGCAACTGTCATTTGAAA	AAAGTTGTTCTT	CTAGTGTCGAAAGGCCCACTGTGT	64515				
194	CTTCCAGTGTAGTGTGTACGA	AGGCGCTTACGACGTGACAGAACTCAC	253					
64514	CTTCCAGTGTAGTGTGTACGA	AGGCGCTTACGACGTGACAGAACTCAC	64455					
254	AGACCCAAAGGTACCGGACAT	GTGTCCCGTGGGTGAGAGGCGCGCATCA	313					
64454	AGACCCAAAGGTACCGGACAT	GTGTCCCGTGGGTGAGAGGCGCGCATCA	64395					
314	ATCAGATGACAGTGTGTCTG	TGGCAGACAGTGTTCAGAAATATTAATAGT	373					
64394	ATCAGATGACAGTGTGTCTG	TGGCAGACAGTGTTCAGAAATATTAATAGT	64335					
374	TTTAAACCTGTAAGCGCGGAC	GACGTGTTCACGCCAGTACAGAAAGCAAGGGA	433					
64334	TTTAAACCTGTAAGCGCGGAC	GACGTGTTCACGCCAGTACAGAAAGCAAGGGA	64275					
434	AGCAGTGTAGCTGATTAAGAA	AGGTTGAAAACAGAAAGCACTTCTACTATCTACAA	493					
64274	AGCAGTGTAGCTGATTAAGAA	AGGTTGAAAACAGAAAGCACTTCTACTATCTACAA	64215					
494	AAAAATCTCCGAATGATTTAT	CAGAAAGATCTTATATGACAGTCAACATATTTGCTGT	553					
64214	AAAAATCTCCGAATGATTTAT	CAGAAAGATCTTATATGACAGTCAACATATTTGCTGT	64155					
554	TAAAGAGGGGTCTTAAGAAAG	CACTGCTTAAGTATTAACATGTGAGAGTGGCAGTT	613					

Db 64154 TAAAGAGGGGCTCTAAGAAAAAGCACTGCTAAGTTAGCACTGTGAGAGTGGCCAGTT 64095
QY 614 TAAATATGACTCAACGCCCATCTGGGAGGAGACAGACAGGGGGAAGGGGCTCAAGAG 673
Db 64094 TAAATATGACTCAACGCCCATCTGGGAGGAGACAGACAGGGGGAAGGGGCTCAAGAG 64035
QY 674 AGACACTGATTAAGATGGCCATTTGTCACTACTGTTTGACAGAAATTAACCGTTAAAA 733
Db 64034 AGACACTGATTAAGATGGCCATTTGTCACTACTGTTTGACAGAAATTAACCGTTAAAA 63975
QY 734 GCTTACCGGTGACCTTTTATTCATGTAATTAACCAATGTAATGTAATGTAATTA 793
Db 63974 GCTTACCGGTGACCTTTTATTCATGTAATTAACCAATGTAATGTAATGTAATTA 63915
QY 794 ATCTTACTTCAATTAATTAAGTCAAAATCTGTCTCTCTCTTGTGATGACGTCTGTTCACA 853
Db 63914 ATCTTACTTCAATTAATTAAGTCAAAATCTGTCTCTCTCTTGTGATGACGTCTGTTCACA 63855
QY 854 CACTCACCACGACACCCACGACGACTAGAAACAGAAATCTCTTAAAGGCAACAGAGAG 913
Db 63854 CACTCACCACGACACCCACGACGACTAGAAACAGAAATCTCTTAAAGGCAACAGAGAG 63795
QY 914 CCAGATCTCTGTTCAAAAGCCCTGAGAAAGCCGCTCAGCTGTAATTTAGAACTCACTAT 973
Db 63794 CCAGATCTCTGTTCAAAAGCCCTGAGAAAGCCGCTCAGCTGTAATTTAGAACTCACTAT 63735
QY 974 GAAATCAAAAGACAGAGCTGTACACCCATCTGTACGTACAGTACAAAGTTACGTAAATGA 1033
Db 63734 GAAATCAAAAGACAGAGCTGTGTACACCCATCTGTACGTACAGTACAAAGTTACGTAAATGA 63675
QY 1034 GCATGGGCTGATTAAGTTACAGGTGGTTACATGAGCGGTGATTAAGGAGGCTGTGCT 1093
Db 63674 GCATGGGCTGATTAAGTTACAGGTGGTTACATGAGCGGTGATTAAGGAGGCTGTGCT 63615
QY 1094 GTGTCAACAGGCTGTGGAGCTACGAGAGGGTCTGCAACCCCTGAGCCAGAAAGCTGTGAGTC 1153
Db 63614 GTGTCAACAGGCTGTGGAGCTACGAGAGGGTCTGCAACCCCTGAGCCAGAAAGCTGTGAGTC 63555
QY 1154 TTCTTAAAGACAAAGCTCTCAACAGCTTAAGTCTTAAGTCTCAACAGCAACGCACT 1213
Db 63554 TTCTTAAAGACAAAGCTCTCAACAGCTTAAGTCTTAAGTCTCAACAGCAACGCACT 63495
QY 1214 TAGTTCACAAGATTTTGGCAATTTCTTAATCTGACAGAAATAGGGGATTTT 1266
Db 63494 TAGTTCACAAGATTTTGGCAATTTCTTAATCTGACAGAAATAGGGGATTTT 63442

RESULT 3
AX163751/c 1017 bp DNA linear PAT 22-JUN-2001
LOCUS AX163751
DEFINITION AX163751
ACCESSION AX163751
VERSION AX163751.1 GI:14544857
KEYWORDS
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1
REFERENCE
AUTHORS Gould-Rothberg, B.E., DiPippo, V.A., Ramsen, T.M. and Gerwehn, R.W.
TITLE Method of identifying toxic agents using nsald-induced differential
JOURNAL gene expression in liver
Patent: WO 0138579-A 15 31-MAY-2001;
Curagen Corporation (US)
FEATURES
source 1.1017
Location/Qualifiers
1.1017
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
BASE COUNT 245 a 245 c 216 g 311 t
ORIGIN

Query Match 80.3%; Score 1017; DB 6; Length 1017;
Best Local Similarity 100.0%; Pred. No. 3.3e-257;
Matches 1017; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 245 GAACCTCAGACAGACCAGGATACCGGAGACATGTGCGCGTGGTGTGAGAGTCTAGAGGG 304
Db 1017 GAACCTCAGACAGACCAGGATACCGGAGACATGTGCGCGTGGTGTGAGAGTCTAGAGGG 958
QY 305 GCGGCTCAATCAGACAGAGCTGTACTCTGGGAAGACAGATGATGTTTCAATATTC 364
Db 957 GCGGCTCAATCAGACAGAGCTGTACTCTGGGAAGACAGATGATGTTTCAATATTC 898
QY 365 TAAATATGTTTAAATACGTAAAGCCGACAGCAGTATTTCTACACCCAGTTACTAGAAA 424
Db 897 TAAATATGTTTAAATACGTAAAGCCGACAGCAGTATTTCTACACCCAGTTACTAGAAA 838
QY 425 ACAGAGGAGACGCTGATGCTGATTAAGGAAAGGTGAAACAGAGAGGCACTTCTACT 484
Db 837 ACAGAGGAGACGCTGATGCTGATTAAGGAAAGGTGAAACAGAGAGGCACTTCTACT 778
QY 485 ATCTACCAAAAAATCTCCGAATGATTAATCAGAAAGATCTTAATAGAGTCAAGAT 544
Db 777 ATCTACCAAAAAATCTCCGAATGATTAATCAGAAAGATCTTAATAGAGTCAAGAT 718
QY 545 ATTGCTCTGTTAAGAGGGGCTCTAAAGAAAGCACTTGTGTAAGTACAGACTGTAGGA 604
Db 717 ATTGCTCTGTTAAGAGGGGCTCTAAAGAAAGCACTTGTGTAAGTACAGACTGTAGGA 658
QY 605 TGCGCAGTTTAAATATGAGACTCAACGCCCATCTGGGAGAGGACAGAGGGGAGGGG 664
Db 657 TGCGCAGTTTAAATATGAGACTCAACGCCCATCTGGGAGAGGACAGAGGGGAGGGG 598
QY 665 GCTCAAGAGACAGATTAAGATGATGCGCCATTTGTCTACTGTTTGACAGAAATTAAC 724
Db 597 GCTCAAGAGACAGATTAAGATGATGCGCCATTTGTCTACTGTTTGACAGAAATTAAC 538
QY 725 CGTTTAAAGCTTTACCCGAGACATTTATAGAGTGAATTTCTCCATGTACAAATGAG 784
Db 537 CGTTTAAAGCTTTACCCGAGACATTTATAGAGTGAATTTCTCCATGTACAAATGAG 478
QY 785 TGTAAATTAATCTCTACTCATATTAATTAATTAATTAATTAATTAATTAATTAATTA 844
Db 477 TGTAAATTAATCTCTACTCATATTAATTAATTAATTAATTAATTAATTAATTAATTA 418
QY 845 TGTTCACACACTCCACCCAGCACACCCAGCAGTACAGAAACAGAAATCTTCTTAAAGCA 904
Db 417 TGTTCACACACTCCACCCAGCACACCCAGCAGTACAGAAACAGAAATCTTCTTAAAGCA 358
QY 905 ACACAGAGGACAGATTTGTTCAAAAGCCGAGAAAGCCGCTGAGCTGATTTTGAAGA 964
Db 357 ACACAGAGGACAGATTTGTTCAAAAGCCGAGAAAGCCGCTGAGCTGATTTTGAAGA 298
QY 965 ACTCAGTATGAATTAAGAGCAGAGCTGTTACACCCATCTGACGTACAGTACAAAGTT 1024
Db 297 ACTCAGTATGAATTAAGAGCAGAGCTGTTACACCCATCTGACGTACAGTACAAAGTT 238
QY 1025 ACCTAATGAGCATGGGCTGATTAAGTTACAGGTGCTTACATGCGACAGCTGTCTTAAGGA 1084
Db 237 ACCTAATGAGCATGGGCTGATTAAGTTACAGGTGCTTACATGCGACAGCTGTCTTAAGGA 178
QY 1085 GGCTGAGCTGTGACAGGCTGAGAGCTGAGGAGGGGCTGACACCCCTGAGCCCAAG 1144
Db 177 GGCTGAGCTGTGACAGGCTGAGAGCTGAGGAGGGGCTGACACCCCTGAGCCCAAG 118
QY 1145 GCTGAGCTGTTTAAAGACAAAGCTCTCAACAGCTTAAGTCTGCTTCTCAAGCA 1204
Db 117 GCTGAGCTGTTTAAAGACAAAGCTCTCAACAGCTTAAGTCTGCTTCTCAAGCA 58
QY 1205 AACGCAACTTATGTTCAACAGGTAATTTTGGCAATTTCTTAATCTGAGCAAGAAATAGGG 1261
Db 57 AACGCAACTTATGTTCAACAGGTAATTTTGGCAATTTCTTAATCTGAGCAAGAAATAGGG 1

RESULT 4
BC003828/c 2319 bp mRNA linear ROD 16-APR-2003
LOCUS MGC:6235 IMAGE:3593957, complete cds.
DEFINITION Mus musculus RAS-related G3 botulinum substrate 1, mRNA (cDNA clone
ACCESSION BC003828 GI:13277917
VERSION BC003828.1
KEYWORDS MGC.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sclerognathi; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 2319)
Straussberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,
Klausner, R.D., Collins, F.S., Wagner, L., Shemen, C.M., Schuler, G.D.,
Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,
Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F.,
Diatelkon, M., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,
Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,
Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S.,
Abramson, R.D., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J.,
McKernan, R.J., Malek, J.A., Gunaratne, P.H., Richards, S.,
Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.M.,
Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,
Fahey, J., Helton, E., Kettelman, M., Madan, A., Rodriguez, S.,
Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y.,
Bouffard, G.G., Blakeley, R.W., Touchman, J.W., Green, E.D.,
Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,
Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smalls, D.E.,
Schermer, A., Schein, J.E., Jones, S.J., and Marra, M.A.
Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
22388257
JOURNAL
MEDLINE
PUBMED
22388257
REFERENCE
1 (bases 1 to 2319)
Straussberg, R.
AUTHORS
TITLE
JOURNAL
DIRECT SUBMISSION
Submitted (28-FEB-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: <http://mgc.ncl.nih.gov>
CONTACT
Email: cgabbs-remail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Sequencing Group at the Stanford Human Genome
Center, Stanford University School of Medicine, Stanford, CA 94305
Web site: <http://www-shgc.stanford.edu>
Contact: (Dickson, Mark) mdickpax11.stanford.edu
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,
R. M.
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAC Plate: 11 Row: e Column: 22.
FEATURES
SOURCE
Location/Qualifiers
1..2319
/organism="Mus musculus"
/mol_type="mRNA"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="MGC:6235 IMAGE:3593957"
/tissue_type="Mammary tumor. Metallothionein-TGF alpha
model. 10 month old virgin mouse. Taken by biopsy."
/clone_lib="NCI CGAP_Mam1"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6"
1..2319
/gene="Rac1"

CDs
/note="synonym: D5Erid559e"
/db_xref="LocusID:19353"
/db_xref="MGI:97845"
198..776
/product="Rac1 protein"
/protein_id="AAH03828.1"
/db_xref="GI:13277918"
/db_xref="LocusID:19353"
/translation="MQAIKCYVVGAGVCKTLLISYTNAPRGYIPIYVDPNCSAN
HYDKPVLNGLMDINGQEDYDRPLSPQIDVFLICSLVSPASFENRKAWEVR
HHCPTPIIVGTCKLIDRDKDPIERLEKELTPITPYDGLAMAREIGAVKYLECSAL
TORGKLTVEDEIRAVILCPYPVKRRKRCILL"
BASE COUNT 603 a 573 c 503 g 640 t
ORIGIN
Query Match 69.4%; Score 878.2; DB 10; Length 2319;
Best local Similarity 85.1%; Pred. No. 1.5e-220;
Matches 1116; Conservative 0; Mismatches 143; Indels 52; Gaps 10;
1 TTTTTTTTTTTTTTTTCAAGTCCAAAGCATTTTTTTTTTTTTTTTATGATTCAGG 60
|||||
2318 TTTTTTTTTTTTTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 2259
61 ATTATTAATGATCATGATCAAAACATAGCTGCTAATGAGCAAAAGATCAATGTAA 120
|||||
2258 ATTATTAATGATCATGATCAAAACATAGCTGCTAATGAGCAAAAGATCAATGTAA 2199
121 AAACACTCCAAATTCGTGCACTGTCATTTGAAAAAGTTGTCTAGTGTCAAGGC 180
|||||
2198 AAACACTCCAAATTCGTGCACTGTCATTTGAAAAAGTTGTCTAGTGTCAAGGC 2139
181 CCAACACTGTGTTCTGCGAGTGAATGTTGTTGTCAGAGGCGTTGCACTAGCGTT 240
|||||
2138 CCAACACTGTGTTCTGCGAGTGAATGTTGTTGTCAGAGGCGTTGCACTAGCGTT 2079
241 GACAGAACCTTCACAGACCCAAAGG-----TACCGAAGCAGTGTG 279
|||||
2078 GACAGAACCTTCACAGACCCAAAGAACCTCTAGGAGAGCCATGACAGAAAGGTGTG 2019
280 TCCGCGTGGGTGAGTCTGAGAGGGGCGGCATCATCATGACAGTGTGTACTCTGG 339
|||||
2018 TCCAGCGTGGCGAGGTCTAGA---GGCAGCATTAATCATGACAGTGTGTACTCTGT 1962
340 CAAGACAGTGAATGTTTCAATATCTAAATATGTTTAAATGTTAAAGCCGACAGCT 399
|||||
1961 AGAAGTTCATATGTTTAAATATCTAAATATGTTTAAATGTTTAAAGCCGACAGCT 1902
400 GATTTCTACACCCAGTCTACTAGAAAACGAAAGGAGCACTAGT-CAGCTGACTAAAGNA 458
|||||
1901 GATTTGACACACTAGTGTGTAAGAACTAGGAACAGCACTAGTCTGAGTAAAGTAA 1842
459 GGTGAAACAGAGACGACTTCTACTATCTACCAAAAAAATCTCCGA-----TGCATTA 513
|||||
1841 GCGGAAACAGAGACGACTTCTACTATCTACCAAAAAAATCTCTCGTTA 1782
514 TCAGAAAGTCTTATAGTACAGTCAAGTCAATCTCTGTTAGAAAGGGCGCTTAAAG 573
|||||
1781 TCAGAAAGTCTTATAGTACAGTCAAGTCAATCTCTGTTAGAAAGGGCGCTTAAAG 1723
574 AAAGCACTGCTAATGTAAGCACTGTGAGATGGCCAGTTTAAATGAGACTCAACGCC 633
|||||
1722 AAAGCACTGCTAATGTAAG-ACCTGAGAGGTGGCCAGTTTAAATGAGACTCAACGCC 1664
633 CATCTGGGAGGAGACAGCAGGGGAGGAGGGGCTCAG-----AGAGCACTG 681
|||||
1663 CATCTGGGAGGAGAGCAGAGTGGAGGAGGGGCGCCAGAGGATGCTCAAGAGCACTG 1604
682 AATAAGTGGCCATTTGTCATCTAGTTTGCACAAATTAACGTTTAAAGCTTACC 741
|||||
1603 AATAAGTGGCCATTTGTCATCTAGTTTGCACAAATTAACGTTTAAAGCTTACC 1544
742 CGTGACACTTATTCAGTGAATTAATCAATGATCAATGATGAATTAATTAATCTTAC 801

|||||
Db 1543 CGGACACTTATTAAGTTGAATTAAGTACTCCATGATACAAATGATGTAAGTAATCTCTAC 1484
QY 802 TTTCATTTAGTCAAAATTAAGTGTCTGTCTCTTTGATGAGTGTGTTTACACATCTCAC 861
Db 1483 TTTCATTTAGTCAAAATTAAGTGTCTGTCTCTTTGATGAGTGTGTTTACACATCTCAC 1426
QY 862 CCAGCACACCCAGCAGTACGAGACAGAAATTAAGTGTGTTAGAGGACACAGGAGCCAGATT 921
Db 1425 CCAGCACACCCAGCAGTACGAGACAGAAATTAAGTGTGTTAGAGGACACAGGAGCCAGATT 1366
QY 922 CTGTTCAAAGCCGAGAAAGCCGATGATGATTTTATAGAGAACTCATATGAATTCAA 981
Db 1365 CTGTTCAAAGCCGAGAAAGCCGATGATGATTTTATAGAGAACTCATATGAATTCAA 1306
QY 982 AGAGCAGAGCTGTTTACACCATC--GTGACGTACAGTACAAAGTTACGTAATGAGCATG 1038
Db 1305 AGAGCAGAGCTGTTTACACCATC--GTGACGTACAGTACAAAGTTACGTAATGAGCATG 1246
QY 1039 GCGTGAATGATTACAGTGTGCTTACATGGCAGGCTGTCATTAAGAGAGCTGTCTGTCTC 1098
Db 1245 GCGTGAATGATTACAGTGTGCTTACATGGCAGGCTGTCATTAAGAGAGCTGTCTGTCTC 1186
QY 1099 ACACGCTGTGGAGCTACGAGGAGGCTGTCTGACACCCGAGCCAGAAAGCTGAGTCTCTT 1158
Db 1185 ACACGCTGTGGAGCTACGAGGAGGCTGTCTGACACCCGAGCCAGAAAGCTGAGTCTCTT 1126
QY 1159 AAGGACAAA--GTCTCTCAACAGCTTAAGTGTCTTACGTTCTTCAGACCAACCACTTA 1215
Db 1125 AAGGACAAA--GTCTCTCAACAGCTTAAGTGTCTTACGTTCTTCAGACCAACCACTTA 1066
QY 1216 GTTCACAGAGTATTTGGCAATCTTAATCTGAGCAAGAAATGAGGAGTTT 1266
Db 1065 GTTCACAGAGTATTTGGCAATCTTAATCTGAGCAAGAAATGAGGAGTTT 1015

RESULT 5
LOCUS BC051053 2281 bp mRNA linear ROD 14-APR-2003
DEFINITION Mus musculus, clone MGC:58966 IMAGE:5038182, mRNA, complete cds.
ACCESSION BC051053
VERSION BC051053.1 GI:29835221
KEYWORDS MGC.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 2281)
Strausberg, R.
Direct Submission
Submitted (11-APR-2003) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: <http://mgc.ncl.nih.gov>
Contact: MGC help desk
Email: cgabs-remail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: National Institutes of Health Intramural
Sequencing Center (NISC),
Gaithersburg, Maryland;
Web site: <http://www.nisc.nih.gov/>
Contact: nisc_mgc@hgrl.nih.gov
Blakesley, N., Ayle, K., Beckstrom-Sternberg, S.M., Benjamin, B.,
Blakesley, R.W., Boufard, G.G., Breen, K., Brinkley, C., Brooks, S.,
Dietrich, N.L., Grante, S., Guan, X., Gupta, J., Hachibhi, P.,
Hansen, N., Ho, S.-L., Karlins, E., Kwong, P., Lalic, P., Legaspi, R.,
Maduro, Q.L., Mastello, C., Maskell, B., Mastrian, S.D., McCloskey, J.C.,
McDowell, J., Pearson, R., Stanlipop, S., Thomas, P.J., Touchman, J.W.,
Tsougen, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L.,
Young, A., Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LNL at: <http://image.lnl.gov>
Series: IRAP Plate: 108 Row: f Column: 9
This clone was selected for full length sequencing because it
passed the following selection criteria: Hexamer frequency ORF
analysis, genomeScan gene prediction, Similarity but not identity
to protein.
FEATURES
Source location/Qualifiers
1..2281
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C2BEH II"
/db_xref="taxon:10090"
/clone="MGC:58966 IMAGE:5038182"
/issue_type="Mammary tumor metastasized to lung. Tumor
arose spontaneously from a senescent normal mammary
(clonal) outgrowth infected with the virus MHV."
/clone_1kb="MGI:CGAP_Lu29"
/lab_host="DH10B"
/note="Vector: pCMV-Sport6"
197..775
/codon_start=1
/product="unknown (protein for MGC:58966)"
/protein_id="A85105.1"
/db_xref="GI:29835222"
/translation="KQAIKCVVGDGAVKTCILISYTNAPPEGEIPIVFDNYSNV
MDGKPVNLGMDTAGEYDRLRPLSPQTDVFLCFSLVSPASENTRAKYKPYR
HMCNPIILVGTGKLDLDDKDIETKEKELPIYPOGLMAKKEIGAVKYLECSAL
TORGKTVGEBAIRAVICPPVKKRKICLL"
BASE COUNT 568 a 569 c 504 g 640 t
ORIGIN
Query Match 68.6%; Score 868.2; DB 10; Length 2281;
Best local Similarity 86.4%; Pred. No. 6.5e-218;
Matches 1102; Conservative 0; Mismatches 123; Indels 50; Gaps 11;
QY 35 TTTTCTTTTCTTTTATGATCAAGATTATTAAGTCATACATGCAAAACATACGCTCA 94
Db 2281 TTTTCTTTTCTTTTATGATCAAGATTATTAAGTCATACATGCAAAACATACGCTCA 2222
QY 95 ACTGATTTAGCAAAAGATCAATGTAAACACATCTCCAAATTTGCAACTGTCAATTGAA 154
Db 2221 ACTGATTTAGCAAAAGATCAATGTAAACACATCTCCAAATTTGCAACTGTCAATTGAA 2162
QY 155 AAGTTTGTCTAGTGTGCGAAAGGCCCAACACTGTGTTCTGCCAGTGAATAGTTGT 214
Db 2161 AAGTTTGTCTAGTGTGCGAAAGGCCCAACACTGTGTTCTGCCAGTGAATAGTTGT 2102
QY 215 ACAGAACGGGCTAGACCTAGCCCTTGACACAACTCAGACACCCAAAG----- 264
Db 2101 ACAGAACGGGCTAGACCTAGCCCTTGACACAACTCAGACACCCAAAGCTCTCA 2042
QY 265 -----TACCGGACATGTGTCGCCGTGGGTGAGAGGTCTAGAGGGGGCGGATCA 313
Db 2041 GCGAGACCCAGACAGAACCGCTGTCTCCACGTCGGGAGAGTGTAG-----GCGACATTTA 1985
QY 314 ATCAATGACAGTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 371
Db 1984 GTCAATGACAGTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1925
QY 372 GTTTAAATCTGTTAAAGCCGACAGCTGATTTCTACACCATTTACTAGAAAAAGG 431
Db 1924 GTTTAAATCTGTTAAAGCCGACAGCTGATTTCTACACCATTTACTAGAAAAAGG 1865
QY 432 GAAGCACTAGT--CAGCTGTAAAGAGGTGAAGAGTGAAGAGGACGCTTCTACTATCTAC 490
Db 1864 AAGCACTAGTCTGATGATTAAGTGAAGGCGAAAGACAGACGACTTCTACTGAGCTAC 1805
QY 491 C-AATAAATCTCGAATGATGATGAGAAAGATCTTATAGTACAGGTGACAGATATTGC 549
Db 1804 C-AATAAATCTCGAATGATGATGAGAAAGATCTTATAGTACAGGTGACAGATATTGC 1745

QY	550	TCGTTAAGAAAGGGGGTCCCTAAAGAAAGACCTTGCTAGTTAGACAACTGTAGAGATGGCC	609
Db	1744	TCGTTAAGAA - GGGGTTCTTAAAGAAAGACCTTGCTAGTTAG - AACTGTAGAGGTTGCC	1687
QY	610	AGTTTAAATATGAGACTAACGCCCATCTGGGGAGGAGACACAGGCGGAAAGGGGGCTCA	669
Db	1686	AGTTTAAATATGAGACTAACGCCCATCTGGGGAGGAGAGCGGCAGTGAGAGGGTGGGGGCG	1627
QY	670	AG-----AGAGACACTGAATAAGATGGCCATTGTGCATCTACTGTGTGACAGA	717
Db	1626	CGAAGGGATGCTCAAGAGACACTGATTAAGATTGGCCATTGTGCATCTACTGTGTGACAGA	1567
QY	718	AATTAACCGTTAAAGAGTTTAAACCGTGACACTTTATTCAGTGAATTAATCTCATGAC	777
Db	1566	AATTAACCGTTAAAGAGTTTAAACCGTGACACTTTATTCAGTGAATTAATCTCATGAC	1507
QY	778	AATGTAGTGTAAATTAATCTCTACTCTCATTTAGTTCAAAATACGTCTGTCTCCGTTGAT	837
Db	1506	AATGTAGTGTAAATTAATCTCTACTCTCATTTAGTTCAAAATACGTCTGTCTCCGTTGAT	1447
QY	838	GACGTCGTGTTCAACACCTCCACCCAGACACACCAGACGACTAGGAACGAATACTTCGT	897
Db	1446	CATGTCCGTGTT--CACACTCCACCCAGACACACCACCACTAGGAACGAATACTTCAT	1389
QY	898	AGAGGCAACACAGNAGCCAGAGTCTGTTCAAAACCTGCAABAACCGGTACAGCTGTAT	957
Db	1388	AGAGGCAACACAGGAAACCCAGAGTCTGTTCAAAACCTGCAAAAACCTACTGCGGTGTT	1329
QY	958	TTAGAGAAGCTCAGTATGAAATCAAGAGACAGCGTGTACACCCTC---GTGAGCTTACA	1014
Db	1328	TTAGAAAAGCTCAGTATGAAATCAAGAGACTAGCTGTACACCCTCATCTACTGTAGCTTACA	1266
QY	1015	GTACAAAGTTACGTATATGAGCATGGGCTGATTAAGTTACAGGTCGCTTACATGCAAGCGTG	1074
Db	1268	GTACAAAGTTACGTATATGACATGGGCTGATTAAGTTACAGGTGCTTACATGCAACGCTG	1209
QY	1075	TCATTAAAGAGGCTGCTGCTGTGTACACAGGCTCTGGAGACTACGGGAGGCTGTGCAACCTT	1134
Db	1208	TCATTAAAGAGAGGCTGCTGCTGTGTACACAGGCTCTGGGAACCTTGAAGGGCTGTGCAACCTT	1149
QY	1135	GAGCCCAAGAGCTCAGTCTTCTTAAAGACAAA--GTCTCTCAACAGACTTATGTCCTTAC	1191
Db	1148	GAGCTCAAGAGACTCAGTCTTCTTATGCAACGAAGTCTCTCAAGAGAGTTATGTCCTTATG	1088
QY	1192	GTGTGTTCTACACAAACGCACTTATAGTTCACAAAGGTATTTTGGCAATTCCTTAATCTGACCA	1251
Db	1088	GTGTGTTCTACAGCAACAAACCTTATGTTTCAGAAAGTATTTTGGCAATTTTCTTAATCTGACCA	1029
QY	1252	AGAATAGGGAATTT 1266	
Db	1028	AGAATGGGGGATTTT 1014	

TITLE
JOURNAL

COMMENT

Iosifidis, I.P., Shim, C., Becker, J., Thomas, E., Perera, A., Gordon, M., Goltz, J.S. and Kucherlapati, R.
Direct Submission
Submitted (03-MAY-2000) Department of Molecular Genetics, Albert
Einstein College of Medicine Genome Center, 1300 Morris Park Ave.
Bronx, NY 10461, USA
On Aug 11, 2001 this sequence version replaced gi:14939654.

```

-----Genome Center
Center: Harvard Partners Genome Center
Center Code: HPGC
Web site: http://www.hpcsg.org/sequence/mouse.html
Contact: hpgc@emdel.mgh.harvard.edu
-----Summary Statistics
Center project name: ABN
Sequencing vector: pUC18; L08752
Chemistry: Dye-terminator Big Dye; 100%
*Consensus quality: 256302 at least Q20
*Consensus quality: 254124 at least Q30
*Consensus quality: 250276 at least Q40
Estimated insert size: agarose-FP - N/A
**Estimated insert size: 268581 - sum-of-ctngs
Quality coverage: agarose-FP - N/A
Quality coverage: 6.2 x In Q20 bases; sum-of-ctngs estimation

```

NOTE: This is a 'working draft' sequence. It currently consists of 26 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of 'N', but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1	51730:	contig of 51730 bp in length
2	51731	51750: gap of unknown length
3	51751	82423: contig of 30673 bp in length
4	82424	82443: gap of unknown length
5	118159	contig of 35715 bp in length
6	118179	118178: gap of unknown length
7	139442	contig of 21263 bp in length
8	139462	139461: gap of unknown length
9	160659	160638: contig of 21177 bp in length
10	160659	160658: gap of unknown length
11	177612	contig of 16553 bp in length
12	177632	177631: gap of unknown length
13	190992	190991: contig of 13360 bp in length
14	191012	191011: gap of unknown length
15	200129	200128: contig of 9117 bp in length
16	200148	gap of unknown length
17	200149	gap of unknown length
18	200149	200148: contig of 11734 bp in length
19	211882	contig of unknown length
20	211883	211902: gap of unknown length
21	211903	220772: contig of 8871 bp in length
22	220774	220773: gap of unknown length
23	220794	228643: contig of 8850 bp in length
24	229664	229663: gap of unknown length
25	229664	229663: contig of 8676 bp in length
26	238340	238339: gap of unknown length
27	238360	244848: contig of 6489 bp in length
28	244868	244867: gap of unknown length
29	244869	249725: contig of 4857 bp in length
30	249726	249725: gap of unknown length
31	249746	255523: contig of 2778 bp in length
32	252524	252523: gap of unknown length
33	252544	257520: contig of 4977 bp in length
34	257521	257540: gap of unknown length
35	257541	260396: contig of 2856 bp in length
36	260397	260416: gap of unknown length
37	260417	262709: contig of 2293 bp in length
38	262710	262729: gap of unknown length
39	262730	265934: contig of 1205 bp in length
40	263935	265934: gap of unknown length
41	263955	264196: contig of 242 bp in length
42	264197	264196: gap of unknown length
43	264217	264216: gap of unknown length
44	264217	264629: contig of 413 bp in length

*	26450	26469	gap of unknown length
*	26450	26508	contig of 959 bp in length
*	265609	265628	gap of unknown length
*	265629	266935	contig of 1307 bp in length
*	266935	266955	gap of unknown length
*	266956	268335	contig of 1380 bp in length
*	268336	268355	gap of unknown length
*	268356	268349	contig of 194 bp in length
*	268550	gap of unknown length	
*	268570	269081	contig of 512 bp in length.

FEATURES

source

```

misc_feature      1. 51730
                   /note="assembly_name:Contig138"
misc_feature      51751. 82423
                   /note="assembly_name:Contig137"
misc_feature      82444. 118158
                   /note="assembly_name:Contig136"
misc_feature      118179. 139441
                   /note="assembly_name:Contig135"
misc_feature      139462. 160638
                   /note="assembly_name:Contig134"
misc_feature      160659. 177611
                   /note="assembly_name:Contig133"
misc_feature      177632. 190991
                   /note="assembly_name:Contig132"
misc_feature      191012. 200128
                   /note="assembly_name:Contig131"
misc_feature      200149. 211862
                   /note="assembly_name:Contig130"
misc_feature      211903. 220773
                   /note="assembly_name:Contig129"
misc_feature      220794. 229643
                   /note="assembly_name:Contig128"
misc_feature      229664. 238339
                   /note="assembly_name:Contig127"
misc_feature      238360. 244848
                   /note="assembly_name:Contig126"
misc_feature      244869. 249725
                   /note="assembly_name:Contig125"
misc_feature      249746. 252523
                   /note="assembly_name:Contig124
                   clone_end:SP6
                   vector_side:left"
misc_feature      252544. 257520
                   /note="assembly_name:Contig123"
misc_feature      257541. 260396
                   /note="assembly_name:Contig122"
misc_feature      260417. 262709
                   /note="assembly_name:Contig121"
misc_feature      262730. 263934
                   /note="assembly_name:Contig120"
misc_feature      263935. 264196
                   /note="assembly_name:Contig119"
misc_feature      264217. 264629
                   /note="assembly_name:Contig118"
misc_feature      264650. 265608
                   /note="assembly_name:Contig117"
misc_feature      265629. 266935
                   /note="assembly_name:Contig116"
misc_feature      266956. 268335
                   /note="assembly_name:Contig115"
misc_feature      268356. 268549
                   /note="assembly_name:Contig114"
misc_feature      268570. 269081
                   /note="assembly_name:Contig113"

```

ORIGIN

Query Match	68.1%;	Score 862;	DB 2;	Length 269081;
Best Local Similarity	84.8%;	Pred. No. 4.3e-216;		
Matches 1101; Conservative	0;	Mismatches 145;	Indels 52;	Gaps 10;

QY	14	TTTTCAAGTCCAAAGACATTTTTTTTTTTTTTTATATTCACAAAGATTTATTAAGTCA	73
Db	63492	TTGCAAAATTCAAAAACAAACGATTAACCGATTGTATATTCAGAGATTTTATTAATCA	63433
QY	74	TACATGCAAAACATACTGCTACATGCGATTAGCAAAAGATCAATGTAAACACCTGCACAA	133
Db	63432	TACATGCAAAACATACTGCTAATCGATTAGCAAAAGATCATTGTAAACACCTGCACAA	63373
QY	134	TTCTGCAACTGCTCAATTGAAAAAGTTTGTCTAGTGTGTCGAAAGGCCACACTGTGTT	193
Db	63372	TTCTGCAACTGCTCAATTGAAAAAGTTTGTCTAGTGTGTCGAAAGGCCACACTGTGATT	63313
QY	194	CTTGCACAGTACAGTTAGGTTGTACAGAACGGCGTTTACACTACCGCTTGACAAACCTCAC	253
Db	63312	CTTGCACAGTACAGTTAGGTTGTACAGAACGGCGTTTACACTACGAGTTGACAAACCTCAC	63253
QY	254	AGACCCAAAG-----TACCGGAAGCATGTGTCCGGTGGGTGA	292
Db	63252	AGACCCCAAGCAATCTTAGCGACAGCCATGACAGAGCGTGTCTCCACGTGGGGGA	63193
QY	293	GGTCTAGAGGGGGGGCGCATCATGACATGACAGTGTGGTACTCTGGCAACAGTGAATG	352
Db	63192	GGTCTAGA---GGCAGCATTAATGACATGACAGTGTGGTACTGTTTAAAGGTTCAATAG	63136
QY	353	TTTTCAGAAATTTCTAAATTAATTTAAAAACGTAAACCCGACAGCTGATTCTACACC	412
Db	63135	TTTTCAGAAATTTCTAAATTAATTTAAAAACGTAAACCCGACACATGATTTGACACACT	63076
QY	413	AGTTACTAGAAAACGAAGGAGACACTACT--CAGCTGAGTTAAAGAAAGGTGAAAAACGA	471
Db	63075	AGTTGCTAGAAAACCTAAGAAAGCACTACAGCTCTGAGTAAAGTAAAGGCAAAAACAGA	63016
QY	472	ACGCACCTCTACATCTTACCAAAAAAATCTCCGA-----TGCATTATGAAAGATCTT	526
Db	63015	CGCACCTCTACTGAGCTACCAAAAAAATCTCTGGTATTCAGAAACATCTT	62956
QY	527	ATTACTACAGGTCACACATATTGCTCGTTAGAAAGGGGTCCTTAAAGAAAGCACTTGCTA	586
Db	62955	ATTACTACAGGTCACACATATTGCTCGTTAGAA--GGGGTTCTTAAAGAAAGCACTTGCTA	62897
QY	587	AGTTAGCAACTGTGAGATGGCCAGTTTAAATTTGCACTCACGCCCATCTTGGGGAGGG	646
Db	62896	AGTTAG--AACTGTGAGGGTGGCCAGTTTAAATTTGCACTCAACGCCCATCTTGGGGAGGG	62838
QY	647	ACAGCAGGGGGAAGGGGGGCTCAAG-----AGAGCACTGATTAAGTCGGGCA	694
Db	62837	ACGGCAGTGGAGGGGTGGGGGGCGGAAGGATGCTCAGAAAGACACTGATTAATGGGCA	62778
QY	695	TTTGTGATCTTACTGTTTGCAGAAATTAACCGTTAAAAAGCTTTACCGGTGACACTTTTA	754
Db	62777	TTTGTGATCTTACTGTTTGCAGAAATTAACCGTTAAAAAGCTTTACCGGTGACACTTTTA	62718
QY	755	TTTCAAGTAAATTAATCATGTACAAATAGTAGTAATTAATCTCTACTCATTAATAGTCA	814
Db	62717	TTTCAAGTAAATTAATCATGTACAAATAGTAGTAATTAATCTCTACTCATTAATAGTCA	62658
QY	815	AAATACTGTCTCTCTCTTTGATGACGTCGTGTTTACACACTCCACCCAGCACCCAC	874
Db	62657	AAATACTGTCTCTCTCTTTGATGATGATGCTGTGTT--CACACTCCACCCAGCACCCAC	62600
QY	875	GACTAGGAAGAAATTAATCGTTAGAGGCAACACAGACGACGATCTTGTTCAAAAGCTT	934
Db	62599	AACTAGGAAGAAATTAATTAATTAAGAGCAACACAGAAACCAAGATCTTGTTCAAAAGCTT	62540
QY	935	GCAGAACCGGTACGCTGATTTTATTAAGAAACCTACTATGAATCAAAAGACAGAGCTGT	994
Db	62539	GCAGAAACGAGTGGCGTGGCTTTTATTAAGAAACCTACTATGAATCAAAAGACGAGCTGT	62480

Db	1473	CTTTGATGCA-----CACACACCCACCCGGGCAACCCACAGGTA--AACGAAATT	1426
Oy	891	CTTCGTTAGAGGCACACAGGAGCCAGAGTTCTGTTCAAAAGCTCGCAGAAAGCCGGTACG	950
Db	1425	CTTTATTTAGAGGAATA----GCAGTTCGTTCAAAATCTCCGCAAAAGACTGTGACGA	1372
Oy	951	TGCTATTTTATAGAACTCCTATGATGAATCAAAAGACAGAGCTGTATACCCATCGTGACG	1010
Db	1371	AAACTCGGTATGAAATTCACAAAGACTGATCCAAAGAGCTGAGTGTCTACGCTCACTCCAT	1312
Oy	1011	TACAGTACAAAGTTACGTTATGAGCATGGGCTGTATAGTTACAGGTGCGTTTACATGSCAG	1070
Db	1311	TACAGTACAAAGTTATGTGCGGGAACAGTGCTGCTTACTCACTGCTGAGTTCATATGCGAA	1252
Oy	1071	CGTCTCAATTAGAGAGCGTGTCTGTGTGCACAGGCTGTGG-----GAGCTACGGGGAGGG	1123
Db	1251	CGTTCATTGGGGAGGCGTGTCTCTCTTATCCGATCTGAGAACTACATAGGAGACCAAGTG	1192
Oy	1124	TCTGCACCC--CTGAGACCCAGAAAGCTGTGACGTCTTCTTAAGGACAAAGTCTCCACAGCTT	1182
Db	1191	TCTGCACCTCTTAAGTGTGAGAAAGTACCGTCTTCTCAAAAGCAGAAAGCTTTTGCMAAGTT	1132
Oy	1183	AGTGTCTACGTGTCTTCAGACAAAC--GCAACTTACTTCACAAAGTATTTTGGCAATTC	1239
Db	1131	CAGTGTCTGCGTGTCTCGGGACACAAACATGACGTAGTTCAGAAAGTATTTTGGCAACTC	1072
Oy	1240	TTAATCTGAGCAAGAAATAGGGG	1261
Db	1071	TTAATCTGACAAAGATAGGGG	1050

RESULT 8					
LOCUS	HSALJ32695/c				
DEFINITION	Homo sapiens rac1 gene.	28567 bp	DNA	linear	PRI 09-NOV-2000
ACCESSION	AJ132695				
VERSION	AJ132695.5				
KEYWORDS	GI:8574037 alternative splicing; Alu; AluYo; AlusG1; AlusP; AlusX; AT-rich; CT-rich; MIR; rac1 gene; rac1 protein; RacB protein; repetitive sequence.				
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
REFERENCE	Macos,P., Skauf,J., Marques,B., Beck,S., Verissimo,F., Gespach,C., Boavida,M.G., Scherer,S.W. and Jordan,P. Small GTPase Rac1: structure, localization, and expression of the human gene Biochem. Biophys. Res. Commun. 277 (3), 741-751 (2000)				
TITLE	JOURNAL MEDLINE 20517245 11062023				
AUTHORS	2 (bases 1 to 28567) Jordan,P.				
TITLE	Direct Submission				
JOURNAL	Submitted (03-FEB-1999) Jordan P., Centro de Genetica Humana, Laboratorio de Oncobiologia, Instituto Nacional de Saude /Dr. Ricardo Jorge, Avenida Padre Cruz, 1649-016 Lisboa, PORTUGAL On Jun 20, 2000 this sequence version replaced g1:7248282.				
COMMENT	Related location A1638561.				
FEATURES	location/Qualifiers				
SOURCE	1..28567 /organism="Homo sapiens" /mol_type="genomic DNA" /db_xref="taxon:9606" /chromosome="7" /map="7p22" /clone="H..NH0425A05, RPCI 11" 1..109 /rpc_family="AlusG" 125..422 /rpc_family="AlusC" 441..742				

[illegible]

[illegible]

Db	27375	TTTTTACCCTGACACTTTTCATTCCTGTCACAAATGAGTGA-----AATCTGCACCTTCGT	27320
Qy	807	A-TTAGACAAATAAATCTGTCTGTCTCTCTTTATATACGTCGGTTTCACACACTCCACCAG	865
Db	27319	ATTTTGTCAAAATATCTCTCTTTGTCTTGTATGA-----CACACACCCACCCGG	27270
Qy	866	CACACCCACGACTAGGAACAGAAATACCTGTTAGAGGCAACACAGAGCCAGACTCTGT	925
Db	27269	CACACCCACAGCTA--AAGAAATTTCTTATTAGAGAAATA-----GCAGTTCTGTTC	27218
Qy	926	TCAAAAGCTTCAGACGCCGCTCAGCTGATTTTATAGAAATCTCAATGAAATCAAGAAG	985
Db	27217	AAAATTCCTCCCAAAAAGTGTCTGAGAAAACTCGGTATGAAATCAAAAAGACTGATCCAAAG	27158
Qy	986	CAGAGCTGTATACCCATCTGACGTACAGTACATACAAAGTTACGTAATGAGCATGGCTGAT	1045
Db	27157	AGCTGACCTGTACAGCTCATCTCATTTACATATACAAATTTATATGCGGAAACGTCGTCT	27099
Qy	1046	AAGTTACAGCTGCTTACATATGACAGCGCTGTCAATTAAGAGAGCTGTCTGTTCACACGGT	1105
Db	27097	AATCTACATGTGATTAATTAATGACCAACCTTCATTCGGGAAGCTGTCTGTTCATCCATAC	27033
Qy	1106	CTGG-----GAGCTACGGGAGGAGGTGTGCACC--CTGAGCCCAAGAGCGCAGTCTTCT	1157
Db	27037	TGACAACTACATFAGGAGACGACAGTGTCTGCACCTCTTAATCTGACAGAGCTACCGTCTCT	26978
Qy	1158	TAAAGACAAAGTCTCTCAACAGCTTAAGTCTTACGCTTCTTCTCACAAC--GCAACTT	1214
Db	26977	CAAAGACGAAAGGTCTTTGCAAAAGTCTAGTCTGCTGCTTCTGCGCAACAAATGACAGTGT	26918
Qy	1215	AGTTCACAAAGTATTTGGCAATTCCTTAATCTGACGACAAAGTATGGGG	1261
Db	26917	AGTTCACAAAGTATTTGGCAATTCCTTAATCTGACAAAGTATGGGG	26871
RESULT 9			
LOCUS	AC009412/c	212827 bp	DNA linear PRI 07-NOV-2001
DEFINITION	Homo sapiens BAC clone RP11-425p5 from 7, complete sequence.		
ACCESSION	AC009412		
VERSION	AC009412.6	GI:14190769	
KEYWORDS	HTG.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
AUTHORS	1 (bases 1 to 212827)		
TITLE	Sulston,J.E. and Waterston,R.		
JOURNAL	Toward a complete human genome sequence		
MEDLINE	Genome Res. 8 (11), 1097-1108 (1998)		
PUBMED	99063792		
REFERENCE	9847074		
AUTHORS	2 (bases 1 to 212827)		
TITLE	Hou,S., Maupin,R., Haakenson,W., Gregory,S. and Belter,E.		
JOURNAL	The sequence of Homo sapiens BAC clone RP11-425p5		
REFERENCE	Unpublished		
AUTHORS	3 (bases 1 to 212827)		
TITLE	Waterston,R.H.		
JOURNAL	Direct Submission		
REFERENCE	Submitted (21-AUG-1999) Genome Sequencing Center, Washington		
AUTHORS	University School of Medicine, 4444 Forest Park Parkway, St. Louis,		
TITLE	MO 63108, USA		
JOURNAL	4 (bases 1 to 212827)		
REFERENCE	Waterston,R.H.		
AUTHORS	Direct Submission		
TITLE	Submitted (23-MAY-2001) Genome Sequencing Center, Washington		
JOURNAL	University School of Medicine, 4444 Forest Park Parkway, St. Louis,		
REFERENCE	MO 63108, USA		
AUTHORS	5 (bases 1 to 212827)		
TITLE	Waterston,R.		
JOURNAL	Direct Submission		
REFERENCE	Submitted (07-NOV-2001) Department of Genetics, Washington		
AUTHORS	University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA		

```

COMMENT
-----
On May 23, 2001 this sequence version replaced gl:134331187.

----- Genome Center

Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: saplens@watsn.wustl.edu
----- Summary Statistics
Center project name: H_NH0425P05
-----

NOTICE: This sequence may not represent the entire insert of this
clone. It may be shorter because we only sequence overlapping
clone sections once, or longer because we provide a small overlap
between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate
chemistry, or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by sequence
from more than one subclone; and the assembly was confirmed by
restriction digest.

MAPPING INFORMATION:
The sequence of this clone was established as part of a mapping and
sequencing collaboration between the NHGRI Chromosome 7 Mapping
Project (Eric D. Green, Director), John D. McPherson in the
Department of Genetics (Washington University), and the Washington
University Genome Sequencing Center. For additional information
about the map position of this sequence, see
http://www.nhgri.nih.gov/DIR/CTB/CHR7, send
mailto:egreen@nhgri.nih.gov, or see http://genome.wustl.edu/gsc

SOURCE INFORMATION:
The RP01-11 human BAC library was made from the blood of one male
donor, as described by Osogawa, K., Moon, P. Y., Zhao, B., Frengen, E.,
Tateno, M., Catanesse, J. J. and de Jong, P. J. (1996) An improved
approach for construction of bacterial artificial chromosome
libraries. Genomics 51:1-8. The clone may be obtained either from
Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong
and coworkers at the Roswell Park Cancer Institute
(http://bacpac.med.buffalo.edu)
VECTOR: pBACe3.6

NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the right is CTD-2195F2, 200 bp overlap.
Actual start of this clone is at base position 1 of RP11-425P5;
actual end is at base position 212633 of RP11-425P5.

Location/Qualifiers
1..212827
    /organism="Homo sapiens"
    /mol_type="genomic DNA"
    /db_xref="taxon:9606"
    /chromosome="7"
    /map="7"
    /clone="RP11-425P5"
    /clone_1b="RPC1-11"
    /length=1696
    /rpt_family="MIR"
    repeat_region
        1698..2006
    /rpt_family="Alu"
    repeat_region
        2024..2337
    /rpt_family="Alu"
    repeat_region
        3006..3093
    /rpt_family="L1"
    repeat_region
        3453..3552
    /rpt_family="L2"
    repeat_region
        4924..5046
    /rpt_family="ERV.L"
    repeat_region
        5318..5693
    /note="similar to EST BF996284 (MID:g12402607)"
    misc_feature
        5580..5845
    repeat_region
        /rpt_family="Alu"
        5925..6277
    repeat_region
        5925..6277

```


misc_feature	/rpt_family="MaLR"	6842..7041	/note="similar to EST BF369507 (NTD:g1331532)"
repeat_region	/rpt_family="Alu"	7364..7661	
repeat_region	/rpt_family="Alu"	8029..8048	
repeat_region	/rpt_family="(CMAA)n"	8379..8678	
misc_feature	/rpt_family="Alu"	8560..8762	
repeat_region	/note="similar to EST AA699398 (NTD:g2702592) z140a03.s1"	8720..9031	
misc_feature	/rpt_family="Alu"	9083..9360	
repeat_region	/note="similar to EST AA633603 (NTD:g2556817) ae66b02.s1"	9598..9899	
misc_feature	/rpt_family="Alu"	9871..10098	
repeat_region	/note="similar to EST AA699398 (NTD:g2702592) z140a03.s1"	10746..11044	
repeat_region	/rpt_family="L1"	11075..11101	
repeat_region	/rpt_family="AT_rich"	11102..11372	
repeat_region	/rpt_family="Alu"	11411..11465	
repeat_region	/rpt_family="GA-rich"	11552..11644	
repeat_region	/rpt_family="L2"	11685..12193	
repeat_region	/rpt_family="L1"	12194..12486	
repeat_region	/rpt_family="Alu"	12487..13003	
repeat_region	/rpt_family="L1"	13004..13305	
repeat_region	/rpt_family="Alu"	13306..13501	
repeat_region	/rpt_family="L1"	13503..13892	
repeat_region	/rpt_family="L1"	13893..14224	
misc_feature	/note="similar to EST BE061102 (NTD:g8405752)"	14244..14544	
repeat_region	/rpt_family="Alu"	14546..14828	
repeat_region	/rpt_family="Alu"	15091..15169	
repeat_region	/rpt_family="MIR"	15672..16449	
repeat_region	/rpt_family="ACHobo"	16640..16844	
repeat_region	/rpt_family="Alu"	16917..16959	
repeat_region	/rpt_family="ACHobo"	16960..17287	
repeat_region	/rpt_family="Alu"	17288..17572	
misc_feature	/rpt_family="ACHobo"	17610..17938	
repeat_region	/note="similar to EST BF996427 (NTD:g12402750)"	17793..18068	
repeat_region	/rpt_family="Alu"	18251..18537	
misc_feature	/rpt_family="Alu"	18485..18864	
misc_feature	/note="similar to EST BE062347 (NTD:g8406997)"	18661..18809	
repeat_region	/note="similar to EST BG186544 (NTD:g13708231)"	18957..19160	
repeat_region	/rpt_family="L1"	19314..19528	
repeat_region	/rpt_family="L1"		

Db	175463	TGATTAATACCAACCTTTCTTTCATCATCTAGCATTTGACAGAAATTAACCTTTAGAG	1754040
Qy	750	TTTTATTTCAGTTGAA---TTACTCCATGACAAATGAGTGTAAATTAATCTACTTCAAT	806
Db	175403	TTTTTACCGTGACACTTTTCATCTCTTGACATATGATGTGA---AAATCTCCACTTTCGT	1753488
Qy	807	A-TTAGTCAAATACTGTCTGTCTCTCTCTTTGATGAGAGCTCTGTCTTCCACACTCCACCAG	865
Db	175347	ATTTTGTCAAAATACTGTCTTTGTCTCTTGATCA-----CACACACCCCAACCCG	1752988
Qy	866	CACACCCACGACTAGSAACAGAAATCTTCTGTTTAGAGGCCAACACAGGACCAAGTTCTGT	925
Db	175297	CACACCCACAGCTA--AACGAAATCTTCATTATGAGGAAATA-----CGAGTTCTGTGT	1752466
Qy	926	TCAAAAGCCTCGACGAGCCGCTCAGCTGGATTTTATGAGAACTCATATGAAATCAAAAG	985
Db	175245	AAAATCTCCGCCAAAAGCTGTGTGAGAAAATCCCTATGAAATCACAAGACTATGCCAAG	1751866
Qy	986	CAGAGCTGTATACACCCATCTGTAGCTAGTACAGTACAAAGTTACGTAAATGAGCATGGGTGAT	1045
Db	175185	AGCTGAGCTGTACCTCCTCACTCCATTACAGTACAAATGTTATGTCCGGAAACACCTGCTGT	1751266
Qy	1046	AAGTTACAGGTCCTGTATACATGGCAGCGTGCATTTAAGAGAGGCTGTCTGTGCACACGGT	1105
Db	175125	AACCTACTGTGTAGTTCAATATGCAACGCTTCAATTCGGGAGGCTGTCTGTCTTACGCATC	1750666
Qy	1106	CTGG-----GAGCTACGGGAGGCTCTGACC-CCTGAGCCCAAGAGCTGCAGTCTTCT	1157
Db	175065	TGAGAACTACATAGGAGAGCAAGTGTCTCACCCTCTTAATCGCAAGACTACCGTCTTCT	1750066
Qy	1158	TAAAGACAAAGTCTGTCAACAGCTTATAGTCTACGTCTCTACAGCAAC--GCNACTT	1214
Db	175005	CAGAAGCAAGGTCTTTTGCAGAAAGTTCAGTCTGGTCTGTCTGGCCACAACAAATGACGTGT	1749466
Qy	1215	AGTTCACAAGGTATTTTGGCAATTCCTTAATCTGAGCAAGATAGGGG	1261
Db	174945	AGTTCAGAAGGTATTTTGGCAACTCTTATCTGAAACAGATAGGGGG	174899
RESULT 10			
AK054993/C			
LOCUS			
DEFINITION	AK054993	2315 bp	mRNA
ACCESSION	AK054993		linear
VERSION	AK054993.1	GI:16549633	
KEYWORDS	Oligo capping; fls (full insert sequence).		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	1 Tashiro, H., Yamazaki, M., Watanabe, K., Kumagai, A., Itakura, S.,		
	Fukuzumi, Y., Fujimori, Y., Komiyama, M., Sugiyama, T., Irie, R.,		
	Otsuki, T., Sato, H., Wakamatsu, A., Ishii, S., Yamamoto, Y., Isono, Y.,		
	Kawai-Hiro, Y., Saito, K., Nishikawa, T., Kimura, K., Yamashita, H.,		
	Matsumu, M., Nakamura, Y., Sekine, M., Kikuchi, H., Kanda, K.,		
	Wagatsuma, M., Murekawa, K., Kanehori, K., Takahashi-Fujii, A.,		
	Oshima, A., Sugiyama, A., Kawakami, B., Suzuki, Y., Sugano, S.,		
	Nagahata, K., Masuno, T., Nagai, K. and Isogai, T.		
	NEDO human cDNA sequencing project		
TITLE	Unpublished		
JOORNAL	2 (bases 1 to 2315)		
REFERENCE	Isogai, T., Otsuki, T. and Sugiyama, T.		
AUTHORS	Direct Submission		
TITLE	Submitted (24-OCT-2001) Takeo Isogai, Helix Research Institute,		
JOORNAL	Genomics Laboratory, 1532-3 Yana, Kiserazu, Chiba 292-0812, Japan		
COMMENT	(E-mail: genomics@helix.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986)		
	NEDO human cDNA sequencing project supported by Ministry of		
	Economy, Trade and Industry of Japan: cDNA full insert sequencing:		
	Research Association for Biotechnology (RAB): cDNA library		
	construction: Helix Research Institute (HRI) (supported by Japan		
	Key Technology Center etc.) ; 5' - & 3' - end one pass sequencing: RAB,		

HRI, and Biotechnology Center, National Institute of Technology and									
Evaluation: clone selection for full insert sequencing: RAB and									
HRI.									
FEATURES									
Source									
Location/Qualifiers									
1..2315									
/organism="Homo sapiens"									
/mol_type="mRNA"									
/db_xref="taxon:9606"									
/clone="BRACE2008968"									
/tissue_type="cerebellum"									
/clone_1id="BRACE2"									
/note="cloning vector: PME18SFlu3"									
BASE COUNT									
603 a 531 c 501 g 680 t									
ORIGIN									
Query Match 34.0%; Score 430.6; DB 9; Length 2315;									
Best Local Similarity 67.8%; Freq. No. 1.8e-102;									
Matches 859; Conservative 0; Mismatches 324; Indels 84; Gaps 15;									
QY	50	ATGATTCAGAGATTATTAAGTCATACATGCAAAAACATACGTCTACTGCAATAGCAAAA	109						
DB	2315	ATGATTCAGAGATTATTAAGTCATACATGCAAAAACATTAATGCAATTGACATTAACAAA	2256						
QY	110	GATCAATGTAAAAACCTCCACCATTTGCACTGTCAATTGAAAAAGTTGTTCTTACT	169						
DB	2255	GATCAATGTAAAAACCTCCACCATTTGCACTGTCAATTGAAAAAGTTGTTCTTACT	2196						
QY	170	GGTCGAAGGCCCAACACTGTGTTCTTCCACAGTAGTAGGTGGTTGACGAAGCGCGTAG	229						
DB	2195	GGCTGGAAGGCTCCACGCTGATTCTCCGACGTAGTTAATTGTACGAACATCTGTCG	2136						
QY	230	CACATAGCCC--TTGACAGACCTCACAGACCCAAAG-----TA	266						
DB	2135	CACATAGCCAGATTACACAGAACCTCAGAACCCAAAGACATCAATAGGCAAAACGACTA	2076						
QY	267	CCGGAAGCATGTGTCGCCGTGGGAGGCTGTAAGGGGGGGGCATTCATCATCATGACAGT	326						
DB	2075	CAGAGAGCGGTGTGTCGCCGTGGGAGGTAAGA--GGTCTAGATTGGTCAAGTGACGT	2017						
QY	337	GTTGGTACTGCGCAAGACAGTGATG-----TTTCAGAAATCTTAA	369						
DB	2016	GTCGTAATCTGGCAAGACAGTGAATGTTAAGAGGTCATAGTTAAGAAATTAATCTAA	1957						
QY	370	TAGTTTAAAACTGTAAAGCCGACACAGTATTTCTACACCCAGTACTTGAAAAAGCA	429						
DB	1956	TATTTTAAAACTGTAAAGCTGTCAACACATATTTTTCACCTAGTTTCTGTGAAGAACTTA	1897						
QY	430	GGAAGCACTAGTCACTGAGTAAAGAAAGTGAAGCAAGACGACCTTCTACTACTA	489						
DB	1896	GGAAGCACTATTATAGTCTGTGAAT-----AAGTACATGGAAGACACTTTTACTAATCG	1842						
QY	490	CCAAAAAATCTCCGATGATCATTTACAGAAAGATCTTAACTACAGGTACAGACATATTGC	549						
DB	1841	ACAAAAAATCTCTTAATGACATTTACAAAAAGATTTTAAATACAG--GAGGCAATATTGC	1783						
QY	550	TCGTTAAGAAAGGGGTCTTAAAGAAACACACTGTGTAAGTAAAGCACTGTAGAGATGGCC	609						
DB	1782	TCAGTCAGAAAGGGTCTTAAAGAAACACACTTACTAAGTTAGGCACTTACACAGAACCAAC	1723						
QY	610	AGTTTAAATATGACTCAAGCCCCATCTGGGGAGGAGCAGACAGGGGAGGGGGGCTCA	669						
DB	1722	AGTTTAAAGATTAATTAATGCCCAATTTGGGAGGAGCATGACAGGTGTAAAGAAAGAA	1663						
QY	670	AGAGAGACATGATTAAGATCCGCCATTTGTATCTACTAGTGTGACAGAAATTAACCGTTA	729						
DB	1662	AAGCTTAAAGAAACCTTCTGTGATTAATCAACCTTCTCTTCATCATCTACTGCAATTGA	1603						
QY	730	AAAAGCTTACCCGAGACATTTTATAGTTGA--TTCATCATGTACATAGTATGTG	786						
DB	1602	CAGAAATTAACCTTTTAAAGTTTTTACCCTGACACTTTCAATCTCTTGTACAAATGTAGTG	1543						
QY	787	TAAATTAATCTCTACTCATTA--TTAGTCAAAATACTGTGTCTCTCTTGTATGACGTGTG	845						

pieces.
 AC102775.3 GI:29164654
 HTG: HTGS PHASE1: HTGS DRAFT; HTGS-FULLTOP.
 KEYWORDS
 SOURCE
 ORGANISM
 Mus musculus (house mouse)
 Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 185245)
 Birren, B., Nusbaum, C. and Lander, E.
 Mus musculus, clone RP23-115C10
 Unpublished
 2 (bases 1 to 185245)
 Birren, B., Linton, L., Nusbaum, C., Lander, E., All, A., Allen, N.,
 Anderson, S., Barna, N., Bastien, V., Boguslavsky, L., Boukhalter, B.,
 Brown, A., Camarata, J., Campolano, A., Chang, J., Chazaro, B.,
 Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A.,
 Cooke, P., Deatellano, K., Dewar, K., Diaz, J.S., Dodge, S., Faro, S.,
 Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gaidyna, S.,
 Glade, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,
 Hagos, B., Heatford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R.,
 Jones, C., Kamat, A., Karatas, A., Kells, C., LaRocque, K.,
 Lamasares, R., Landers, T., Lehoczy, J., Levine, R., Liu, G.,
 Maclean, C., MacDonald, P., Major, J., Marquis, N., Matthews, C.,
 McCarthy, M., McEwan, P., McKernan, K., McPheters, R., Meldrum, J.,
 Menus, L., Mihova, T., Mienga, V., Murphy, T., Naylor, J., Nguyen, C.,
 Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D.,
 Olyver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V.,
 Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P.,
 Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupack, R.,
 Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
 Strauss, N., Subramanian, A., Talamas, J., Testafaye, S., Theodore, J.,
 Toplam, K., Travers, M., Travis, N., Triggillo, J., Vassiliev, H.,
 Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G.,
 Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
 Direct Submission
 Submitted (23-NOV-2001) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 3 (bases 1 to 185245)
 Birren, B., Nusbaum, C., Lander, E., Abouelell, A., Allen, N.,
 Anderson, S., Arachchi, H.M., Barna, N., Bastien, V., Bloom, T.,
 Boguslavsky, L., Boukhalter, B., Camarata, J., Chang, J., Choepel, Y.,
 Collymore, A., Cook, A., Cooke, P., Corum, B., Deatellano, K.,
 Diaz, J.S., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Faro, S.,
 Ferreira, P., FitzGerald, M., Gage, D., Galagan, J., Gaidyna, S.,
 Graham, L., Grand-Pierre, N., Hafez, N., Hagopian, D., Hagos, B.,
 Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,
 Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R.,
 Lindblad-Toh, K., Liu, G., Lui, A., Mabbitt, R., Maclean, C.,
 MacDonald, P., Major, J., Manning, J., Matthews, C., McCarthy, M.,
 Meldrum, J., Menus, L., Mihova, T., Mienga, V., Murphy, T., Naylor, J.,
 Nguyen, C., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P.,
 O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N.,
 Rachupara, A., Ramsamy, V., Raymond, C., Retta, R., Rise, C., Rogov, P.,
 Roman, J., Schauer, S., Schupack, R., Seaman, S., Severy, P., Smith, C.,
 Spencer, B., Stange-Thomann, N., Stojanovic, N., Stubbs, M.,
 Talamas, J., Testafaye, S., Theodore, J., Toplam, K., Travers, B.,
 Vassiliev, H., Venkataraman, V.S., Viel, R., Vo, A., Wilson, B., Wu, X.,
 Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
 Direct Submission
 Submitted (23-MAR-2003) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Mar 23, 2003 this sequence version replaced gi:22381790.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RW/RepeatMasker.html
 ----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIRB
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: IL9647

Center clone name: 115_C_10
 ----- Summary Statistics
 Sequencing vector: Plasmid; n/a; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.960731
 Consensus quality: 183758 bases at least Q40
 Consensus quality: 184136 bases at least Q30
 Consensus quality: 184291 bases at least Q20
 Insert size: 178000; agarose-fp
 Quality coverage: 12.6 in Q20 bases; agarose-fp
 Quality coverage: 12.2 in Q20 bases; sum-of-coverage

 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 9 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 1
 53693: contig of 53693 bp in length
 53793: gap of 100 bp
 56775: contig of 2862 bp in length
 56776: gap of 100 bp
 60589: contig of 3814 bp in length
 60590: gap of 100 bp
 67210: contig of 6521 bp in length
 67211: gap of 100 bp
 82585: contig of 15275 bp in length
 82586: gap of 100 bp
 82586: gap of 100 bp
 103119: contig of 20433 bp in length
 103219: gap of 100 bp
 132145: contig of 28227 bp in length
 132146: gap of 100 bp
 132246: gap of 100 bp
 182335: contig of 50090 bp in length
 182336: gap of 100 bp
 182436: contig of 2810 bp in length.
 Location/Qualifiers
 1. 185245
 /organism="Mus musculus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10090"
 /clone="RP23-115C10"
 /clone_1lb="RPCT-23 Female Mouse BAC"
 1. 53693
 /note="assembly-fragment
 clone_end:SP6
 vector_side:left"
 53794. .56675
 /note="assembly-fragment"
 56776. 60589
 /note="assembly-fragment"
 60590. 67210
 /note="assembly-fragment"
 67311. 82585
 /note="assembly-fragment"
 82686. 103118
 /note="assembly-fragment"
 103219. 132145
 /note="assembly-fragment"
 132246. 182335
 /note="assembly-fragment"
 182436. 185245
 /note="assembly-fragment
 clone_end:T7
 vector_side:right"
 BASE COUNT 56870 a 34956 c 35069 g 57550 t 800 others
 ORIGIN
 Query Match 32.2% Score 408.2; DB 2; Length 185245;
 Best Local Similarity 89.8%; Pred. No. 2.2e-96;
 Matches 486; Conservative 0; Mismatches 43; Indels 12; Gaps 4;

DB 184513 T 184513

RESULT 15
AC105979

LOCUS
DEFINITION Mus musculus clone RP24-301E12, WORKING DRAFT SEQUENCE, 21 ordered pieces.

AC105979 192498 bp DNA linear HTG 21-OCT-2002
AC105979
AC105979.4 GI:24182089
HTG: HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
Mus musculus (house mouse)
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.

REFERENCE
AUTHORS Birren, B., Nusbaum, C. and Lander, E.
TITLE
JOURNAL
REFERENCE
AUTHORS 2 (bases 1 to 192498)
Unpublished
2 (bases 1 to 192498)
Birren, B., Linton, L., Nusbaum, C., Lander, E., All, A., Allen, N., Anderson, S., Barna, N., Bastien, V., Boguslavsky, L., Boukhgalter, B., Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S., Farrell, P., FitzHugh, W., Gage, D., Galagan, J., Gardys, S., Glende, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N., Hagos, B., Heatford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamal, A., Karatas, A., Kells, C., Lacroque, K., Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Matthews, C., Maclean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M., McKernan, P., McKernan, K., McPheeters, R., Meldrum, J., Menus, L., Mihova, T., Mienga, V., Murphy, T., Naylor, J., Nguyen, C., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupback, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, K., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Triggillo, J., Vassiliev, G., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE
JOURNAL
REFERENCE
AUTHORS 3 (bases 1 to 192498)
Submitted (11-JAN-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 192498)
Birren, B., Nusbaum, C., Lander, E., All, A., Allen, N., Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavsky, L., Boukhgalter, B., Camarata, J., Chang, J., Chazaro, B., Choepel, Y., Collymore, A., Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S., Farrell, P., FitzGerald, M., Gage, D., Galagan, J., Gardys, S., Gord, S., Graham, L., Grand-Pierre, N., Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamal, A., Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, G., Maclean, C., Macdonald, P., Major, J., Matthews, C., McCarthy, M., Meldrum, J., Menus, L., Mihova, T., Mienga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Raymond, C., Retta, R., Rise, C., Rogov, P., Roman, J., Roy, A., Schauer, S., Schupback, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, G., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE
JOURNAL
COMMENT
Submitted (21-OCT-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On Oct 21, 2002 this sequence version replaced gi:2019424.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: MIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L20199
Center clone name: 301_E.12
----- Summary Statistics
Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 190070 bases at least Q40
Consensus quality: 191243 bases at least Q30
Consensus quality: 191514 bases at least Q20
Insert size: 163000; agarose-fp
Insert size: 191698; sum-of-contigs
Quality coverage: 11.6 in Q20 bases; sum-of-contigs
Quality coverage: 9.9 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 9 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
1
1062 1061: contig of 1061 bp in length
* 1162 5714: contig of 4553 bp in length
* 5715 5814: gap of 100 bp
* 5815 11422: contig of 5608 bp in length
* 11423 11522: gap of 100 bp
* 11523 17600: contig of 6078 bp in length
* 17601 30157: contig of 12457 bp in length
* 30158 30257: gap of 100 bp
* 30258 38787: contig of 8530 bp in length
* 38788 38887: gap of 100 bp
* 38888 99444: contig of 60357 bp in length
* 99445 99445: gap of 100 bp
* 99445 143263: contig of 43719 bp in length
* 143264 143363: gap of 100 bp
* 143364 192498: contig of 49135 bp in length.
FEATURES
source
1..192498
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/clone="RP24-301E12"
/clone_11b="RP24-301E12"
1..1061
/note="assembly-fragment"
misc_feature
1162..5714
/note="assembly-fragment"
misc_feature
5815..11422
/note="assembly-fragment"
misc_feature
11523..17600
/note="assembly-fragment"
misc_feature
17701..30157
/note="assembly-fragment"
misc_feature
30258..38787
/note="assembly-fragment"
misc_feature
38888..99444
/note="assembly-fragment"
misc_feature
99545..143263
/note="assembly-fragment"
misc_feature
143364..192498
/note="assembly-fragment"
vector_end:17
vector_end:right"
BASE COUNT 53596 a 42471 c 43252 g 52379 t 800 others

XX WPI: 2001-355948/37.
 DR Screening hepatotoxic agent comprises contacting test cell population
 XX expressing RISKMARKER or INJURYMARKER with agent, comparing expression
 PT with reference population and identifying difference in expression
 PT levels -
 XX Disclosure: Page 7-8; 76pp; English.
 PS
 XX
 CC The present invention describes a method of screening a test agent for
 CC hepatotoxicity. The method comprises: (a) providing a test cell
 CC population comprising a cell capable of expressing one or more nucleic
 CC acid sequences selected from the group consisting of RISKMARKER 1-8
 CC and INJURYMARKER 1-10; (b) contacting the test cell population with a
 CC test agent; (c) measuring expression of one or more of the nucleic
 CC acid sequences in the test cell population; (d) comparing the
 CC expression of the nucleic acid sequence in the test cell population to
 CC the expression of the nucleic acid sequence in an reference cell
 CC population comprising at least one cell whose exposure status to a
 CC hepatotoxic agent is known; and (e) identifying a difference in
 CC expression levels of the RISKMARKER or INJURYMARKER sequences, if
 CC present, in the test cell population and reference cell population.
 CC The method is useful for identifying a hepatotoxic agent. The present
 CC sequence is given in the exemplification of the present invention.
 CC
 XX
 S0 Sequence 1017 BP; 245 A; 245 C; 216 G; 311 T; 0 other;
 Query Match 80.3%; Score 1017; DB 22; Length 1017;
 Best Local Similarity 100.0%; Pred. No. 1e-244;
 Matches 1017; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 245 GAACCTCAGACACCCAAAGTAGCCGAGCATGTGTCCGCGTGGGTAGTCTAGAGGGG 304
 DB 1017 GAACCTCAGACACCCAAAGTAGCCGAGCATGTGTCCGCGTGGGTAGTCTAGAGGGG 958
 QY 305 GCGGCATCATCATGACATGCTGTGTCTCTGCGCAAGACATGATGTTTCAGAAATATC 364
 DB 957 GCGGCATCATCATGACATGCTGTGTCTCTGCGCAAGACATGATGTTTCAGAAATATC 898
 QY 365 TAAATATGTTTAAATTAAGCTGTAAGCCGAGCATGATTTTACACCCAGTTACTAGAAA 424
 DB 897 TAAATATGTTTAAATTAAGCTGTAAGCCGAGCATGATTTTACACCCAGTTACTAGAAA 838
 QY 425 ACGAAGGGAACACATGCTACGCTAGTAAAGGAAGTGAAGAACAGACGACCTTCTACT 484
 DB 837 ACGAAGGGAACACATGCTACGCTAGTAAAGGAAGTGAAGAACAGACGACCTTCTACT 778
 QY 485 ATCTACCAAAAAATCTCCGAATGCATTATCAGAAAGATCTTATAGTACAGGTACAGCAT 544
 DB 777 ATCTACCAAAAAATCTCCGAATGCATTATCAGAAAGATCTTATAGTACAGGTACAGCAT 718
 QY 545 ATTCGCTGTTAAGAAAGGGGCTCTTAAGAAAGACCTTGTCTAAGTTAGCACTGTGAGGA 604
 DB 717 ATTCGCTGTTAAGAAAGGGGCTCTTAAGAAAGACCTTGTCTAAGTTAGCACTGTGAGGA 658
 QY 605 TGGCCAGTTTAAATATGAGACTCAAGCCCATCTGGGGAGGAGACAGCGGGGAAGGGGG 664
 DB 657 TGGCCAGTTTAAATATGAGACTCAAGCCCATCTGGGGAGGAGACAGCGGGGAAGGGGG 598
 QY 665 GCTCAAGAGACACTGATTAAGATGCGCATTTGTCTACTGTGTTGACAGAAATTAAC 724
 DB 597 GCTCAAGAGACACTGATTAAGATGCGCATTTGTCTACTGTGTTGACAGAAATTAAC 538
 QY 725 CGTTAAAGGCTTAAACCGTGACACTTTTATTCAGTTGAATTAATCTCAATGATAGAG 784
 DB 537 CGTTAAAGGCTTAAACCGTGACACTTTTATTCAGTTGAATTAATCTCAATGATAGAG 478
 QY 785 TGTAAATTAATCTCTACTTATTAATTAAGTCAAAATAGTGTCTGCTCTTATGACGTCG 844
 DB 477 TGTAAATTAATCTCTACTTATTAATTAAGTCAAAATAGTGTCTGCTCTTATGACGTCG 418
 QY 845 TGTTCACACACTCCACCAGACACCCAGACTAGGAAGAAATTAATCTGTTAGAGGCA 904

DB 417 TGTTCACACACTCCACCAGACACCCAGACTAGGAAGAAATTAATCTGTTAGAGGCA 358
 QY 905 ACACAGAGCCAGACTTCTGTTCAAAAGCCTGAGAAAGCCGCTGACCTGTTATTAGAGA 964
 DB 357 ACACAGAGCCAGACTTCTGTTCAAAAGCCTGAGAAAGCCGCTGACCTGTTATTAGAGA 298
 QY 965 ACTCACTATGAATCAAGAGACAGAGCTGTTACACCCATCTGACGTACAGTACAAAGTT 1024
 DB 297 ACTCACTATGAATCAAGAGACAGAGCTGTTACACCCATCTGACGTACAGTACAAAGTT 238
 QY 1025 ACGTATATGACATGGGCTGATATAGTTACAGGTGCGTTACATAGGACGCTGCTATTAGGA 1084
 DB 237 ACGTATATGACATGGGCTGATATAGTTACAGGTGCGTTACATAGGACGCTGCTATTAGGA 178
 QY 1085 GCGTGTGCTGTGTCACAGGCTGCGAGCTACAGGAGGCTGACCCCTAGACCCAGAA 1144
 DB 177 GCGTGTGCTGTGTCACAGGCTGCGAGCTACAGGAGGCTGACCCCTAGACCCAGAA 118
 QY 1145 GCTGAGCTTCTTAAAGCAAAAGTCTCAACACACTTAACTGCTTACGTTCTCAGCAC 1204
 DB 117 GCTGAGCTTCTTAAAGCAAAAGTCTCAACACACTTAACTGCTTACGTTCTCAGCAC 58
 QY 1205 AACGCACTTAGTTCACAAAGGTATTTGGCAATTTCTTAATCTGACCAAGAAATAGGGG 1261
 DB 57 AACGCACTTAGTTCACAAAGGTATTTGGCAATTTCTTAATCTGACCAAGAAATAGGGG 1
 RESULT 3
 AAF18236/C
 ID AAF18236 standard; DNA; 2051 BP.
 XX
 AC AAF18236;
 XX
 DT 14-MAR-2001 (first entry)
 XX
 DE Lung cancer associated polynucleotide sequence SEQ ID 255.
 XX
 KW Human; Lung cancer associated protein; neuroprotective; cytostatic;
 KW cardiostatic; immunomodulatory; muscular active; vulnerable;
 KW gastrointestinal; nephrotoxic; anti-infective; gynecological;
 KW antibacterial; diagnosis; neural disorder; immune disorder; reproductive;
 KW proliferative disorder; wound healing; infectious disease; ds.
 XX
 OS Homo sapiens.
 OS
 PN WO20005180-A2.
 XX
 PD 21-SEP-2000.
 XX
 PF 08-MAR-2000; 2000WO-0505918.
 XX
 PR 12-MAR-1999; 99US-0124270.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 PA (ROSE/) ROSEN C. A.
 PI Ruben SM;
 XX
 DR WPI: 2000-587514/55.
 DR P-PSDB: AAB58360.
 PT Lung cancer associated gene sequences, referred to as lung cancer
 PT antigens, useful for treatment, prevention, and diagnosis of disorders
 PT such as lung cancer -
 XX
 PS Claim 1; Page 716-717; 1425pp; English.
 CC Polynucleotide sequences AAF17982 - AAF18424 encode human lung cancer
 CC associated proteins represented in AAB58106 - AAB58548. Lung cancer
 CC associated proteins and polynucleotide sequences, their agonists, and
 CC antagonists may have neuroprotective; cytostatic; cardiostatic;
 CC immunomodulatory; muscular active general; vulnerable; gastrointestinal

XX		Gene; liver cancer; ds; hepatocellular carcinoma; hepatotropic;
KW		metastatic liver tumour; cytostatic; expression profile; disease state;
XN		disease progression; drug toxicity; drug efficacy; drug metabolism.
XX		Homo sapiens.
OS		WO200229103-A2.
PX		11-APR-2002.
PD		02-OCT-2001; 2001WO-US30589.
PF		02-OCT-2000; 2000US-237054P.
PR		(GENE-) GENE LOGIC INC.
PA		Horne D, Alvares C, Peres-Da-Silva S, Vockley JG;
PI		WPt; 2002-426119/45.
DR		Diagnostics and detecting the progression of liver cancer,
PT		hepatocellular carcinoma or metastatic liver tumor in a patient,
PP		involves detecting the level of expression of two or more genes in a
PS		liver tissue sample -
XX		Claim 1; SEQ ID NO 1632; 298pp; English.
CC		The invention relates to a novel method for diagnosing and detecting the
CC		progression of liver cancer, hepatocellular carcinoma or metastatic liver
CC		tumour in a patient, and differentiating metastatic liver cancer from
CC		hepatocellular carcinoma in a patient, involving detecting the level of
CC		expression of two or more genes represented in ABN93503-ABN97455 in a
CC		tissue sample. The method of the invention has hepatotrophic, and
CC		cystostatic activity. The method is useful for diagnosing and detecting
CC		the progression of liver cancer, hepatocellular carcinoma and metastatic
CC		liver carcinoma in a patient. The method is useful for identifying
CC		expression profiles which serve as useful diagnostic markers as well as
CC		markers that can be used to monitor disease states, disease progression,
CC		drug toxicity, drug efficacy and drug metabolism.
CC		Note: The sequence data for this patent did not form part of the printed
CC		specification, but was obtained in electronic format directly from WIPO
CC		at ftp.wipo.int/pub/published_pct_sequences.
XX		SQ Sequence 1232 BP; 329 A; 240 C; 251 G; 411 T; 1 other;
QY		Query Match 33.2%; Score 420.2; DB 24; Length 1232;
Bd		Best Local Similarity 67.5% ; Pred. No. 3.3e-95;
Matches	848:	Conservative 1; Mismatches 324; Indels 84; Gaps 15
OY	53	ATTCAAGATTTTAAAGTCATACATGCMAAACAATCATGTCTTAAGCATTAGCAAAGAAT 112 :
Dd	1228	WTCCAAGATTTATTAAGTCATACATGCMAAACAATAAATGAATTCATTAGCAAAAAGAT 116 :
OY	113	CATATGAAAAAACCTGCCACAATTCTGCACACTCAATTGAAAAAAAAAGTTTGTTAGTGGT 172
Dd	1168	CAATGTAAAAAACACTGCCACAATTCTGCACACTCAATTGAAAAAAATTCGTGTAGTGCGC 110
OY	173	GGAAGAGCCCCAAGACTGTGTTCTTCCCAGTGTAGTTAGTTGNAGCAAGGGGTTAGAGAC 232
Dd	1108	TGAAAGGCTCCACACGCTGTATTCTCGCCAGTGAGTTGATGTTACAGAAACATGCTGCACC 104
OY	233	TAGCGC--TTGACAGAAACCTCACAGACCCCAAG-----TACCG 269
Dd	1048	TAGCACAATTTACAGAAACCTCACAGACCCCAAGAAACAATCATATGAGCAAAAGGACATACAG 989
OY	270	GAAACATGTCTCCGCTGGGAGTCTTAGAGAGGCGGCAATCATACATACAGACTGTT 329
Dd	988	GAGCAGCTGTCTCCGCTGGGAGTCTTAGAGAGGCGGCAATCATACATACAGACTGTC 930
OY	330	GCTACTCTGGGAAAGCAGTGAAT-----TTTGAANAATCTAAATAAG 372
Dd	929	GCTAATCTGGCAAGCAGTGAATTTTTAAGAGCTCATAGTCTTAAAGAAATCTAAATAAT 870

OY		373	TTTTAAAACTGTAAAGCCGACAGCAGTATTTCTACACCAGTTACTGTGAAGAACAAGG	432
Db		869	TTTTAAAACHTATAAAGCTGCACACATATTTTTCACCTGATTCTTGAAAACCTAAGA	810
OY		433	AAGCACTAGTCAGCTGAGTAAGAAGGTGAAAAAGCAAGCACACTTCTACTATCTACCA	492
Db		809	AAGCACTATTAGCTGTGAATA-----AAGTAACATGTGAAGAACACTTTTACTAATCGACA	755
OY		493	AAAAATCTCCGATNGCATTTATCGAAAAGATCTTATAGTAGTCAGACATATTGGCTCG	552
Db		754	AAAAAACCTTGTATGCAATTCAGAAAAGATTTTATATATACAAAG-GAGGCCATTATGGTCA	696
OY		553	TTAAGAAGGGGGTCCCTAAAAGAAAGCACTTGCTTAAGTTAGCACTGTGAGATGGCCAGT	612
Db		695	GTCGAAAGGGGTTCTATATAGAAAAGACATTACTAATGATGGCACTTACAGAACACAGT	636
OY		613	TTTAATATGAGCTCAACGCCCCCATCTGGGGAGGAGACAGCAGGGGAAAGGGGGGCTCAAGA	672
Db		635	TTTAAGATGATTTAAATGCCAATTTGGGAGGCGATGGCAGGTGTAAAGAAAAGAAAGAA	576
OY		673	GAGCACTGATTAAGATGGCCATTGTCATCTACTGTTTGACAGAAATTAACGTTAAA	732
Db		575	CTTAAGAAAACATTTTCCGTGATTAATACCAACCTTCTCTTCATCACTACTGCATTGTGACG	516
OY		733	AGCTTTACCCGTGACACTTTTATTCAGTTGAA---TTACCTCATGTACAAATGTAGTAA	789
Db		515	AAATTTAACCTTTTAGAGTTTATTCACCGTACACCTTTCTTCTGTTACAAATGATGTGA-	457
OY		790	ATTATATCTCTACTTCATA-TTAGTCAAATATCTGTGTCTCCCTTGTGACGTCGTGT	848
Db		456	---AATGCCACCTTCGTATTATTTGTGCAAAATCTGCTTGTGTCCTTTGATCA-----	409
OY		849	TCAACACACTCCACCAGACACCCACGACACTGAGAACAGATATCTTCGTTAAGCAACAC	908
Db		408	-CACACACCCCACCCCGCACACCCACACACTTA--AACGAAATTCATTTAGAGAAATAG	352
OY		909	AGGAGCGAGATGTCGTTCAAAAGCCTGCAGAAAGCCGGCCAGCTGATTTTAAAGAACTC	968
Db		351	-----CAGTCTGTCTCAAAATCTCCGCAAAAGCTGGTCAGAAAACTCGCTATGAATCA	298
OY		969	ACTATGAATCAAAAGACAGAGCTGTTAACACCCTCGTAGCTACAGTACAAAGTTACGT	1028
Db		297	CAAAGACTGATCCAAGAGCTGAGACTGCTGACGTCACATCCATTCACATGAATGTATGT	238
OY		1029	AATAGCATGGGCTGATATAGTTAACAGTGGCTTACATGGCAGCGCTGTACTTAAAGAGCT	1088
Db		237	CGGGAACACGCGTCGTCACTCACTGTGATTCATAATGGCAACCTTCATTCGGGAAGCT	178
OY		1089	GTGCTGTGTGCACAGGCTGTGG-----GAGCTACGGAGAGGTGTGCACC-CTGAGGCC	1140
Db		177	GTTCGTGCTTTCGCAITCTGAGAACTACATAGAGAACAGCAAGTGTGTGCACCTCTTACTGC	118
OY		1141	AGAAGCTGCACTCTCTTAAAGAGCAAAAGTCTCTACACAGCTTAACTGCTTACGTTCTCA	1200
Db		117	AGAAGCTACCGCTCTTCAAAAGAGCAAGGTCCTTGTCAAGTTCAGTGTGCGTGTCTCG	58
OY		1201	GCACAAC---GCACTTAGTTCAACAAGGATTTTGGCAATCTTAAATCTTGAGCAGA	1254
Db		57	GCACAACAATGCAATGTAGTTCAAGAAGTATTTTGGCAACTCTTAAATCTGAMAACAGA	1
<hr/>				
RESULT 7				
AAH22399/c				
ID	AAH22399 standard; DNA; 348 BP.			
XX	AAH22399;			
XX	AC			
XX	22-AUG-2001 (first entry)			
DE	Human rac1 gene related nucleotide sequence #3.			
XX				
XX	Identification; toxic; hepatotoxic; differential gene expression;			
FW				

Dp	1254	CAATGTAAAAAACTCCACATTTCTGGACACTGTCAATTTTAAAAAATCTGTTGTAGTGGC	1195
OY	173	CGAAAGGCCCAACACTGTGTTCTTGGCAGTAGTTAGGTGTACAGAAAGCGGTTAGCAC	232
Dp	1194	TGAAGGGTCCACCCACTGTATTTCTCGCAGTAGTTAAGTTGTACAGAAACATCGTACAGC	11355
OY	233	TAGGCG--TTGACGAACCTCCAGACCCTAAAGG-----TACCG	269
Dp	1134	TAGACAGGTTTACGAACCTCACAGACCCAAAGAACATCAATAGGCAAGCACTACAG	1075
OY	270	GAACCATGTGTCCCGGTGGTGTGAGAGTGTAGAGGGGGGGCATCAATACATAGTGT	329
Dp	1074	GAGCGGTGTGTCCGCGTGGGGAGAGTAAAGA--GGGTCAGTATTGTGTCAAGTACAGTCTC	1016
OY	330	GGTACTGTGGCAGACAGTATGTTCA-----GAATATCTAAATAG	372
Dp	1015	GGTATCTGTGGCAACAGTAGTATGTTAAGAGGTTCAATAGTTTAAGAAATTAATCTAAATAT	956
OY	373	TTTAAACACGTAAAGCGGCAGCACGATTTCTACACCCAGTTACTATAGAAACGAAGG	432
Dp	955	TTTAAAACTATTAAGGTGCAACACATGATTTTACACTGTACTACAGAAACCTAAGA	896
OY	433	AAGCACTAGTCACTGAGTAAAGAGAGGTGAAAAACAGAACGCACTTCTATCTACCA	492
Dp	895	AAGCACTTATTAGTCTGAAATAA-----GTAACTAGGAAGCACTTTACTAATCGACA	841
OY	493	AAAAATCTCCGAATGATTAATACAGAAAGATCTTATAGTACAGGTGACACATATTGCTG	552
Dp	840	AAAAACCTTCTATGATTAATTCAGAAAGATTTTATTAATCAAG--GAGGCATATTGCTCA	782
OY	553	TTTAAGAGGGGTCTTAAAGAA-----AGCACTTGCTAAGTTAGCACTGTGAGATGC	607
Dp	781	GTCTGAAGGGGTCTTTTAGGAAAGCACACTTACTAAGTTAAGCGCACTAACGAACAA	722
OY	608	CCAGTTTAAATATGAGACTAA---CGGCCCATCTGGGGAGGAGACAGGGGGGAAGGGGG	664
Dp	721	CCAGTTTAAAGCATGAATTAATATCCCAATTTGGGGAGGACATGCGAGGTGTAAAGCA	662
OY	665	GCTCAAGAGA-----GACACTGATTAAGATGGGCATTTGTCTAT	702
Dp	661	AAGGAAAAAGGCTTAAGAAAAACATTTCTTGATTAATCAACAACTTTCTTCATCACTCA	602
OY	703	CTACTGTTTGCACAAATTAACCTTTAAAAAGCTTTACCCGTGACACTTTTATTCAGTTG	762
Dp	601	CTGATTTGAGCAACAAATTAACCTTTTAGAGTTTAAACCCGTGACCACTTTCATATCC	542
OY	763	AATTAATCCATGTACATATGTATTAATTAATCTACTCATATTAATGTCAAAATATCTG	822
Dp	541	CTTGTACCAATGTAGTAGTGTAAATCTCCCACTCCGTAATTTTGTGCCA---AAATTAATG	485
OY	823	TCTGTCTCTTGTAGTAGCTGTGTTTACACACTCCACCCAGACACCCACGACTAGGA	882
Dp	484	TCTTTGTCTTTGA-----TCCACACACACACCCCGGACACGCCACAGCTATAA	430
OY	883	ACAGAAATCTTCTGTAGAGCACACAGAGCCAGAGTCTGTCTAAAGCCTGCAGAGC	942
Dp	429	CCAAATTTCTTCTTAGAGGAAATATACCAAGTTCTGTTCAAAATCTCCCGAAAGCTGG	370
OY	943	CGGTCACTGGTATTTTAGAAGACTCTACTATGAATCAAAAGAGAGAGCTTTACACCCA	1002
Dp	369	TCAGAAAACTCGCTATGAATAATCAAAAGACTGATCAAAAGAGCTGTAGAGCTTACGCTCA	310
OY	1003	TCGTGACGTACATGACAAAGTTAGTATAGATAGATGGGCTG--ATTAAGTTACAGGTGCGTT	1061
Dp	309	--CTTCAATTAAGTACAAATGTATTGTGGGAGAACAGCTGTCTCGTAACTCACTGGTGAATT	252
OY	1062	ACATGCAAGCGTGTATTAAGAGAGCTGTGTGTGTCAACAGGTCTGGGA-----GC	1113
Dp	251	CAATGGCAAGCCTTCATTCGGAGAGCTGTCTGTCTTAACGATCTGAGAACTACATAGGA	192
OY	1114	TAGGGGAGGGGTCTGACC--CCTGAGCCCAAGAAAGTGTGATTTCTTAA--GGACAAAGTC	1170
Dp	191	GAGCAAAAGTGTCTCACCTCTTAATCTGACGAAGGTACCGTCTTCTCAAAATAGCAAGGCT	132

OY	1171	TCTCAACGCGTTAGGGC--TACGGTCCTCCAGACAACGCATCTTGTTCAAGAAGTAT	1228
DB	131	TTTGCAAAAGTTCAGGGCTCGGTGTTCTCGCACAACTATGATGTTCAAGAAGTAT	72
OY	1229	TTTTGCCAATTCCTTAATCTGCACGAATAATAGCGG	1261
DB	71	TTTGCAACTCTTAATCTGTAACAAAGAATGGGG	39
 RESULT 9 AAH23398/c ID AAH23398 standard; DNA; 1022 BP. XX XX AAH23398; AC XX XX DT 22-AUG-2001 (first entry) XX DE Human rac1 gene related nucleotide sequence #2. XX KW Identification; toxic; hepatotoxic; differential gene expression; KM NSAD; non-steroidal antiinflammatory drug; ds. XX OS Homo sapiens. XX PM MO200138579-A2. PD XX 31-MAY-2001. PF 21-NOV-2000; 2000WO-US32049. XX PR 22-NOV-1999; 99US-0166923. PR 18-FEB-2000; 2000US-0183531. PR 20-NOV-2000; 2000US-0717321. XX PA (CURA-) CURAGEN CORP. XX PI Gould-Rothberg BE, DiIppio VA, Ramseh TM, Gerweln RM; XX WPT: 2001-355948/37. DR XX XX Screening hepatotoxic agent comprises contacting test cell population PT expressing RISKMARKER or INJURYMARKER with agent, comparing expression PT with reference population and identifying difference in expression PT levels - PS Disclosure: Page 7-8; 76pp; English.			
XX	The present invention describes a method of screening a test agent for		
CC	hepatotoxicity. The method comprises: (a) providing a test cell		
CC	population comprising a cell capable of expressing one or more nucleic		
CC	acid sequences selected from the group consisting of RISKMARKER 1-8		
CC	and INJURYMARKER 1-10; (b) contacting the test cell population with a		
CC	test agent; (c) measuring expression of one or more of the nucleic		
CC	acid sequences in the test cell population; (d) comparing the		
CC	expression of the nucleic acid sequence in the test cell population to		
CC	the expression of the nucleic acid sequence in an reference cell		
CC	population comprising at least one cell whose exposure status to a		
CC	hepatotoxic agent is known; and (e) identifying a difference in		
CC	expression levels of the RISKMARKER or INJURYMARKER sequences, if		
CC	present, in the test cell population and reference cell population.		
CC	The method is useful for identifying a hepatotoxic agent. The present		
CC	sequence is given in the exemplification of the present invention.		
XX			
SQ	Sequence 1022 BP; 273 A; 208 C; 205 G; 336 T; 0 other;		
Query Match	23.0%; Score 290.6; DB 22; Length 1022;		
Best Local Similarity	64.5%; Pred. No. 8,8e-63;		
Matches	664; Conservative 0; Mismatches 304; Indels 61; Gaps 13;		
OY	265	TACCAGAACATGTGTCCGCGTGGGTGAGCTCTAGAGGGGGCGCATCATCATGACACA	324
DB	1000	TACAGAGAGCGCTGTGTCCGCTGGCGCGAGGTAAAGA-GGGTCACATATTGGTCAGTGACA	942

QY	325	GTGTTGGTACTGCGCAAGACAGATGTTTCACAA-----TATCTAA	367
Db	941	GTGTCGGTATCTGCGCAAGACAGATGTTTAAGAAAGTTCTATTATTAAGATTATCTAA	882
QY	368	AATAGTTTAAAAACTGTAAAGCCGACAGCGTGAATTTCTACACCAGTTACTAGAAAAAG	427
Db	881	AATATTTTTTAAAAAGCATATTAAGCTGCACACATCATGATTTTTTACACCTAGTTACTAGAAAAGT	822
QY	428	AAGGGAAGCACTAGTCAGCTAGTAAAGGAAGGTGMAAACAGGAAGCACTTCTACTATTC	487
Db	821	AAGGAAGACACTTATTAAGCTCTGAAATA---AAGTAACATGAGAAAGCACTTTTACTATTAAT	767
QY	488	TACCAAAAAAATCTCCGAATGCATTATTCAGAAATCTTTATAGTACAGAGTCAGACATATT	547
Db	766	CGACAAAAAACCCTTCTATATGCATTATTCAGAAAGATTTTTATATATATCAAG--GAGCGATATTT	708
QY	548	GCTCGTTTAAAGAGGGGGCTCTTAAGAAAAAGCACTTGTCTAAGTTAGCAACTGTGAGATGG	607
Db	707	GCTCAGTCAGAAAGGGGTTCTTAATAGAAAAAGCACTTACTTAAGTTAGCGCACTTAACGAACAA	648
QY	608	CCAGTTTAAATATGACTCAACAGCCCACTCTGGGAGGGAACAGCAAGGGGGAGGGGGCT	667
Db	647	CCAGTTTAAATGAAATTAATTAATGCCCCAATTTGGGAGGAGCAAGGAGGTGAAGAAAGG	588
QY	668	CAAGAGAGCACTATATAGATCGGCCAATTTGTCACTCTACTGTTGACAGAAATTAACCGT	727
Db	587	AAAAGCTTAAAAAAAATTTCTCTATATATACCAACCTTCTTCTATCATCTACTGCAATTT	528
QY	728	TAAAAAGCTTTACCCGTGACACCTTTTATTCAGTTGAATTT---ACTCCATGTACAATGTAG	784
Db	527	GACAGAAATTAACCTTTATAGAGTTTATACCCGTGACACTTTCAATTCCTTGTACAAATGTAG	468
QY	785	TGTAAATTAATCTCTACTTCAATA--TTGTGCAAAATACGTGCTGCTCTCTTGAAGAGCTC	843
Db	467	TGTA---AATCTCCACTTGTGTAATTTGTCTAAAAATACGTGTTGTGCTCTTGTGATCA---	416
QY	844	GTGTTTCCACACTCCACCCAGCAGCACCCAGCAGTATAGGAACAGATACTTGTTAGAGGC	903
Db	415	-----CACACACCCACCCGGGACACCCAGCAAGCTA--AACGAATTTCTTATTAGAGGA	364
QY	904	AACACAGAGCCAGATTTCTGTTCAAAAGCCTGCAGAAAGCCGGTCACTGTTATTTTAGAG	963
Db	363	AATA-----GCAGTTCTGTTCAAAATCTCCGCAAAAGCTGTGAGAAAACTCGCTATGTA	310
QY	964	AATCTACTATGAAATCAAAAGAGAGAGCTGTTAACACCCTAGTGTGACGTACGTACAAAGT	1023
Db	309	AATCACAAAAGACTGATCCAAAGAGCTGAGCTGTACGCTCACTCCATTAACGTACAAAGT	250
QY	1024	TACGTAATGAGCATAGGCGCTGATTAAGTTTCAGGTGCTGTACATGCGAGCGGTGTCAATAG	1083
Db	249	TATGTCGGGAACAGGTGCTGCTAACTCACTGCTGATGAGTTCATGCGAAGCGCTTCAATTTGGGG	190
QY	1084	AGGCTGTCTGTGTACACAGGCTGTG-----GAGTACGGGAGGCGTCTGCACC--CCTG	1135
Db	189	AGGCTGTCTGTCTTACCATCTGTAGAACTACATATAGGAGACAAAGTGTCTCACCTCCTTA	130
QY	1136	AGCCAGAGAAGTGCAGTCTTCTTAAGCAAGTCTCCAACAGGTTATGTGCTTACGTACGTGT	1195
Db	129	ACTGCAGAGGCTACCGTCTTCTCAAAAGCAGAAAGCTTTTGCAAAAGTTAGTGTCTGCGGTGT	70
QY	1196	TCTGACGACAAC---GCAACTTAGTTTCAAGAGTATTTTGGCAATTTCTTAATCTGAGCAA	1252
Db	69	TCTGCGCACACAAATGACAGTGTAGTTTCAAGAAAGTATTTTGGCACTCTTAATCTGAAACAA	10
QY	1253	GAATTAGGGG	1261
Db	9	GAATGGGGG	1

XX	AA89693;
AC	
XX	
DT	08-JAN-2001 (first entry)
XX	
DE	Mouse Exo103 nucleotide sequence.
XX	
KW	Exocytosis pathway protein; Exo protein; antiparkinsonian; antidiabetic;
KW	antiallergic; antiasthmatic; neuroprotective; anticonvulsant;
KW	vulnerary; asthma; inflammation; allergy; Chediak-Higashi syndrome; CHS;
KW	Alzheimer's disease; Parkinson's disease; Huntington's disease; diabetes;
XX	digestion disorder; wound healing disorder; gene therapy; ss.
OS	Mus sp.
XX	
PN	MO200043419-A2.
XX	
PD	27-JUL-2000.
XX	
PF	20-JAN-2000; 200OWO-US01431.
XX	
PR	20-JAN-1999; 99US-0116534.
PR	26-JAN-1999; 99US-0117272.
PR	26-JAN-1999; 99US-0117308.
PR	26-JAN-1999; 99US-0117309.
PR	26-JAN-1999; 99US-0117312.
PR	01-FEB-1999; 99US-0118177.
PR	01-FEB-1999; 99US-0118178.
PR	01-FEB-1999; 99US-0118179.
PR	09-FEB-1999; 99US-0119286.
PR	11-FEB-1999; 99US-0119999.
PR	11-FEB-1999; 99US-0119759.
PA	(RIGE-) RIGEL PHARM INC.
XX	
PI	Luo Y;
XX	
WP:	2000-482908/42.
PT	New nucleic acids encoding Exo proteins which are useful in the
PT	diagnosis, treatment or prevention of exocytosis-mediated disorders
PT	such as asthma, inflammation and allergies -
XX	
PS	Disclosure: Page 271; 305pp; English.
XX	
CC	The present sequence encodes a polypeptide which is associated with
CC	the exocytosis pathway. cDNA molecules encoding proteins involved in
CC	exocytosis have been isolated by yeast one-hybrid and two-hybrid
CC	screening. Novel proteins, termed Exo proteins, have been identified that
CC	interact with known exocytosis-associated proteins such as GS27, alpha
CC	snap, unc18-l, vsmpt, snap-23, and the rab family of proteins.
CC	Exo proteins and their agonists and antagonists are useful in the
CC	diagnosis, treatment or prevention of exocytosis-mediated disorders
CC	such as asthma, inflammation, allergies, Chediak-Higashi Syndrome
CC	(CHS), Alzheimer's disease, Parkinson's disease, Huntington's disease,
CC	diabetes, digestion disorders and wound healing disorders.
CC	The nucleic acids, antagonists or agonists of Exo proteins are useful
CC	in gene therapy. The nucleic acids are also useful for generating
CC	transgenic or knock-out animals which can be used in the
CC	development and screening of therapeutically useful reagents.
XX	
SQ	Sequence 447 BP; 102 A; 95 C; 93 G; 157 T; 0 other;
Query Match	18.2%; Score 230.6; DB 21; Length 447;
Best Local Similarity	78.4%; Pred. No. 7e-48;
Matches 349; Conservative	0; Mismatches 69; Indels 27; Gaps 5;
OY	77 ATGCAAAATPACTCTCTAATGCATTAGCAAAAAGATGCATGTAAACAATCACAATTC 136
DB	447 ATGCAAAACCTTCTCTTAATCGATTAGCAAAAAGATCATGT-AAAACTTCACAATTC 389
OY	137 TGCAACTGTCAATTTGAAAAGTTTGTC-TAGTGTGGAAGGCCCAACACTGTGTCCT 195

```

Db      388  TGCACACTGCAATTGAAAAAGTTGTTCTTAGTGTTGAAAGCCCAACTGCATTCT 329
Oy      196  TGCACAGTAGTGGTGTGTACAGAGCGCGTTAGCAGTACGCTTACAGAGAACTTCACAG 255
Db      328  TGCCAGTAGTGGTGTGTACAGAGCGCGTTAGCAGTACGCTTACAGAGAACTTCACAG 269
Oy      256  ACCCAAG-----TACCGGAAGCATGTGTCCGCGTGGTGAG 294
Db      268  ACCCAAGCAATCTTAGCAGAGCCATGACAGAGCCGCTGTCCACGTGGCGAGG 209
Oy      295  TCTAGAGGGGCGGCAATGACATGACAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 354
Db      208  TCTAAG---GGCAGATTAGTACAGTACAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 152
Oy      355  TCAGAAATCTAATAATAGTTTAAACCTGTAAAGCCGAGCAGTGTGTGTGTGTGTGTGT 414
Db      151  TAAGAAATCTAATAATAGTTTAAACCTGTAAAGCTGTAAAGCTGTAAAGCTGTGTGTGT 92
Oy      415  TTACTAGAAAGCAAGGAGGAGCACTAGT-CAGCTGAGTAAAGGAGTGTAAAGCAAGAAC 473
Db      91  TTGCTAGCAAACTAAGGAGGAGCACTAAGCTGTGTGTGTGTGTGTGTGTGTGTGTGT 32
Oy      474  GCACCTTCTACTATCTACCAAAAAA 498
Db      31  CACTTCTACTAGTACCAAAAAA 7

```

RESULT 11

AAA89694/c
ID AAA89694 standard; cDNA; 422 BP.

AAA89694;

08-JAN-2001 (first entry)

Mouse Rab2 nucleotide sequence #1.

Exocytosis pathway protein; Exo protein; antiparkinsonian; antidiabetic;
antiallergic; antispasmodic; nootropic; neuroprotective; anticonvulsant;
vulnerary; asthma; inflammation; allergy; Chediak-Higashi syndrome; CRS;
Alzheimer's disease; Parkinson's disease; Huntington's disease; diabetes;
digestion disorder; wound healing disorder; gene therapy; ss.

Mus sp.

WO200043419-A2.

27-JUL-2000.

20-JAN-2000; 2000WO-US01431.

20-JAN-1999; 99US-0116534.
26-JAN-1999; 99US-0117274.
26-JAN-1999; 99US-0117308.
26-JAN-1999; 99US-0117309.
01-FEB-1999; 99US-0117312.
01-FEB-1999; 99US-0118177.
01-FEB-1999; 99US-0118178.
01-FEB-1999; 99US-0118179.
09-FEB-1999; 99US-0119286.
11-FEB-1999; 99US-0119288.
11-FEB-1999; 99US-0119759.

(RIGEL-) RIGEL PHARM INC.

Luo Y.

WPI: 2000-482908/42.

New nucleic acids encoding Exo proteins which are useful in the
diagnosis, treatment or prevention of exocytosis-mediated disorders
such as asthma, inflammation and allergies -

PS Disclosure: Page 271-272; 305pp; English.

XX The present sequence encodes a polypeptide which is associated with
CC the exocytosis pathway. cDNA molecules encoding proteins involved in
CC exocytosis have been isolated by yeast one-hybrid and two-hybrid
CC screening. Novel proteins, termed Exo proteins, have been identified that
CC interact with known exocytosis-associated proteins such as GS27, alpha
CC snap, unc18-1, vamps, snap-23, and the rab family of proteins.
CC Exo proteins and their agonists and antagonists are useful in the
CC diagnosis, treatment or prevention of exocytosis-mediated disorders
CC such as asthma, inflammation, allergies, Chediak-Higashi Syndrome
CC (CHS), Alzheimer's disease, Parkinson's disease, Huntington's disease,
CC diabetes, digestion disorders and wound healing disorders.
CC The nucleic acids, antagonists or agonists of Exo proteins are useful
CC in gene therapy. The nucleic acids are also useful for generating
CC transgenic or knock-out animals which can be used in the
CC development and screening of therapeutically useful reagents.

XX Sequence 422 BP; 95 A; 92 C; 83 G; 148 T; 4 other;

Query Match 17.0%; Score 215.8; DB 21; Length 422;

Best Local Similarity 77.0%; Pred No. 3.5e-44;

Matches 322; Conservative 0; Mismatches 70; Indels 26; Gaps 4;

```

Oy      104  GCAAAAGATCAATGTAAAAAAGT-CCACAAATCTGCAACTGTCAATGTAAAAAGTTTG 162
Db      421  GCAAAAGATCAATGTAAAAAAGT-CCACAAATCTGCAACTGTCAATGTAAAAAGTTTG 362
Oy      163  TTCTAGTGTGTGAAAGGCGCCCAACTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 222
Db      361  TTCTAGTGTGTGAAAGGCGCCCAACTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 302
Oy      223  GCGTTAGCAGTACGCGCTGACAGAACCTACAGACCCCAAGS----- 264
Db      301  GCGTTAGCAGTACGCGCTGACAGAACCTACAGACCCCAAGS----- 242
Oy      265  ---TACCGGAAGCATGTGTCCGCGTGGTGTAGAGGGGCGGCGCATCATCATG 321
Db      241  CATGACAGGAAGCGTGTGTCCACGTCGCGGAGGAGTGTAG---GCGACGATTTAGTCACATG 185
Oy      322  ACAGTGTGTGTCTGTGTGCAAGACAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 381
Db      184  ACAGTGTGTGTCTGTGTGAAAGGTTTCATAGTGTGTGTGTGTGTGTGTGTGTGTGTGT 125
Oy      382  TGTAAAGCGGAGCAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 441
Db      124  GGTAAAGCGGAGCAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 65
Oy      442  T-CAGCTGAGTAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 498
Db      64  TAGCTCTGAGTAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 7

```

RESULT 12

ABL38065
ID ABL38065 standard; cDNA; 605 BP.

ABL38065;

08-APR-2002 (first entry)

Human colon tumour antigen polynucleotide seq ID NO:1654.

Human; colon cancer; colon tumour antigen; cytostatic; vaccine;
colon tumour metastatic antigen; diagnosis; gene; ss.

Homo sapiens.

WO200196388-A2.

20-DEC-2001.

08-JUN-2001; 2001WO-US18557.

QY	368	AAATGTTTAAAACTGTAAAGCCGAGACCGATTTCTTACACCAGTACTATAAAACG	427		
Db	193	AAATTTTAAAAAATTAAGAGCTGTACACATGATTTTACACCTAGTACTAGAAAAC	252		
QY	428	AAGGAGACACTAGTCAGCTGATGAAGAGAGGTGAAAAACAGAACGCATTCTATTC	487		
Db	253	AAGGAAAGCACTATTATGCTGTGAATA-----AAGTACATGAAAGACATTTTACTAT	307		
QY	488	TACCAAAAAAATCTCCGATTCATATTCAGAAAGATCTTATATGATACAGTACATAT	547		
Db	308	CGACAAAAAAACCTTCTTAATGCAATTAACGAAAGATTTTATTAATACAAAG-GAGGCATAT	366		
QY	548	GCATGTTAAAGAGGGGGCTCTTAAGAAAAAGCACTGTCTAAGTTAGACACTGTGAGATG	607		
Db	367	GCCTACGTCAGAAAGGGTTCTTATATAGAAAAAGCACTTACTAATGTTAGCACTAACGAA	428		
QY	608	CCAGTTTAAATATGACTCAAGCCCATCTGTGGAGGAGGACAGAGGGGGAAGGGGCT	667		
Db	427	CCAGTTTAAAGATGATTAATGATGATTTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGG	488		
QY	668	CAGAGAGACACTGATTAAGATGGCCATTTTGC-ATCTACGTGTTGACAGAAATTAACG	728		
Db	487	GAAAGACNTAAGAAACATTTCTGTTATATCAAAACCTTTTNTTATCTACTGNAT	548		
QY	727	TTTAAAGCTTTACCCGTGACACTTTTATTCAGTGAATTCCTCATGTAC	777		
Db	547	TGACAAAAATTAACCTTTTAAAGTTTACCCAGCACTTNTTCTGTTCTC	597		
RESULT 15					
AC	ABX42447/c	standard; cdna; 432 BP.			
XX	ABX42447;				
XX	ABX42447;				
XX	20-FEB-2003 (first entry)				
DE	Bovine EST associated with lactation/muscle/fat deposition #7612.				
XX					
KW	Bovine; ss; EST; expressed sequence tag; lactation; LMFD.				
KW	muscle deposition; fat deposition; genome mapping; gene identification;				
XX	gene analysis; cattle breeding.				
OS	Bos Taurus.				
XX					
PN	US2002137139-A1.				
XX					
PD	26-SEP-2002.				
XX					
PF	24-SEP-2001; 2001US-0960352.				
XX					
PR	12-JAN-1999; 99US-115707P.				
PR	11-JAN-2000; 2000US-0480902.				
XX					
PA	(BYAT/) BYATT J C.				
PA	(MATH/) MATHIALAGAN N.				
PA	(TAON/) TAO N.				
PA	(WARR/) WARREN W C.				
XX					
PI	Byatt JC, Mathialagan N, Tao N, Warren WC;				
XX					
DR	WPL; 2003-110599/10.				
XX					
PT	New nucleic acid associated with lactation, and muscle and fat				
PT	deposition, useful for genome mapping, gene identification and				
PT	analysis, cattle breeding, or for genetically improving cattle				
XX					
PS	Claim 2; SEQ ID No 7612; 245bp; English.				
XX					
CC	The invention relates to a purified nucleic acid molecule associated with				
CC	lactation or muscle and fat deposition (designated LMFD), derived				
CC	from cattle, and the LMFD nucleic acid can specifically hybridise to a				

CC second nucleic acid molecule comprising any of 15112 nucleotide
CC sequences, appearing as ABX34836-ABX49947, or complements of them.
CC Also included are: (1) a transformed cell having a nucleic acid
CC comprising an LMF nucleic acid linked to a promoter and a 3' non-
CC translated sequence that functions in the cell to cause termination of
CC transcription and addition of polyadenylated ribonucleotides to a 3' end
CC of the mRNA molecule; and (2) determining a level or pattern of a
CC molecule in a bovine cell or tissue comprising: (a) incubating a marker
CC nucleic acid (comprising any of the 15112 nucleic acid sequences or its
CC complement or fragment) with a complementary nucleic acid molecule
CC obtained from the bovine cell or tissue, where hybridisation between the
CC marker nucleic acid and the complementary nucleic acid permits the
CC detection of the molecule; and (b) detecting the level or pattern of the
CC complementary nucleic acid, where the detection of the complementary
CC nucleic acid is predictive of the level or pattern of the molecule.
CC The LMF nucleic acid is used for determining a level or pattern
CC of a molecule in a bovine cell or tissue. It is useful for genome
CC mapping, gene identification and analysis, cattle breeding, preparation
CC of constructs for use in cattle gene expression, or for genetically
CC improving cattle. The present sequence is one of the 15112 bovine
CC LMF EST (expressed sequence tag) nucleic acids.
CC Note: The present sequence was not shown in the specification but
CC was obtained in electronic format from the USPTO web site:
CC seqdata.uspto.gov/sequence.html?DocID=20020137139.
XX

Sequence 432 BP: 123 A; 76 C; 78 G; 155 T; 0 other;

Query Match 13.6%; Score 172.2; DB 25; Length 432;
Best Local Similarity 72.2%; Pred. No. 2.9e-33;
Matches 294; Conservative 0; Mismatches 103; Indels 10; Gaps 5;

QY 379 AACTGTAAGCCGACGACGTGATTCTA-CACCCAGTTACTAGAAAACGAGGAAAGCA 437
DB 432 AACCATTAACCTGACGACGTGATTCTTAACCTAGTACTAGAAAACGAGGAAAGCA 373
QY 438 CTAGTCAGC--TGAGTAAAGGAAGTGAAGAAAGCAAGCACTTCTACTACCAAAA 495
DB 372 CTTATTAGCTTTGATTAACCAACATGAAAACAGAGGTGACATTCTACTACCAAAA 313
QY 496 AAATCTCCGATGATTAATGAGAAAGATCTTATAGTACAGGTGACACATATGCTCGTTA 555
DB 312 AAATTTCTAATGATTAATGAGAAAGATTTATTAATACAG-GAGGCATATGCTCATTA 254
QY 556 AGAAGGGGTCCTAAAGAAAGCACTGCTAAGTACAGTACAGTACAGTACAGTACAGT 615
DB 253 AGAATGAGTTCTAATGAGAAAGCACTTACTAAGTACCACTAATGAGATGACCAAGTCA 194
QY 616 AATATGACACTCAAGGCC---CCATCTGGGAGGAGGACAGGAGGGAAGGGGCTCAAG 671
DB 193 GAGATGATTAATGAGGCAATTTCAAGAGGGGTGGGCACTTTAAGGAAAGCTTAAGAAA 134
QY 672 AGAGACACTGATAGATCGGCCATTTGTCATCTACTG--TTTGACAGAAATTAACCGTTA 729
DB 133 ACACCTACCTGATATATACCGACCTTCTTCATCTACCGCATTTGACAGAAATTAACCTTT 74
QY 730 AAAAGCTTTACCGTGCATTTTATTCATTCATTCATTCATTCATTCATTCATTCATTCAT 776
DB 73 AAAACTTTTACCGCTGATGCTTTTAATGTTTAAGTTTACGTTA 27

Search completed: August 25, 2003, 23:47:50
Job time : 355.341 secs

REFERENCE AUTHORS	3 Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Kono, H., Akiyama, J., Nishi, K., Kitsuami, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujisake, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Matsubara, K., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, T., Okazaki, Y., Muramatsu, M., Inoue, Y., Kita, A. and Hayashizaki, Y.
TITLE	RIKEN Integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer
JOURNAL MEDLINE PUBMED REFERENCE AUTHORS	Genome Res. 10 (11), 1757-1771 (2000) 20530913 11076861 4 Kawai, J., Shingawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y., Aizawa, T., Hara, A., Fukunishi, Y., Kono, H., Adachi, J., Fukuda, S., Arakawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamamoto, R., Saito, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Saito, R., Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T., Fleischmann, W., Gaasterland, T., Glasl, C., King, B., Kochiwa, H., Kuehl, P., Lewis, S., Matsuo, Y., Nikaido, I., Pesole, G., Quackenbush, J., Schiraldi, L., Staudt, F., Suzuki, R., Tomita, M., Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H., Balderelli, R., Barash, G., Blake, J., Boffelli, D., Bojunga, N., Carninci, P., de Bonaldo, M., Brownstein, M. J., Bult, C., Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D., Hofmann, M., Hume, D. A., Kamiya, M., Lee, N. H., Lyons, P., Marchionni, L., Mashima, J., Mazzerelli, J., Mombauts, P., Nordone, P., Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H., Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K. F., Suzuki, H., Toyooka, K., Wang, K. H., Weitz, C., Whitaker, C., Winding, L., Wynshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawai, H., Kohzuki, S. and Hayashizaki, Y.
TITLE	Functional annotation of a full-length mouse cDNA collection
JOURNAL MEDLINE PUBMED REFERENCE AUTHORS	Nature 409 (6821), 685-690 (2001) 21085660 11217851 5 The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
TITLE	Analysis of the mouse transcriptome based on functional annotation
JOURNAL REFERENCE AUTHORS	Nature 420, 563-573 (2002) 6 (bases 1 to 2269) Aochi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hirooka, T., Hirozane, T., Horii, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Kono, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, Y., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohnato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shingawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeaki, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.
TITLE	Direct Submission
JOURNAL	Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gscc.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)
COMMENT	cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
FEATURES	Please visit our web site for further details. URL: http://genome.gsc.riken.go.jp/ URL: http://fantom.gsc.riken.go.jp/ Location/Qualifiers

source	1. .2269 /organism="Mus musculus" /mol_type="mRNA" /strain="C57BL/6J" /db_xref="FANTOM:DB:9430013G17" /db_xref="taxon:10090" /clone="9430013G17" /issue_type="embryonic body between diaphragm region and neck" /clone_1lb="RIKEN full-length enriched mouse cDNA library" /dev_stage="12 days embryo" 198, 776
CDS	/note="unnamed protein product; RAS-related G3 botulinum substrate 1 (MSD) MGI:97845, GB AK011072, evidence: BLASTN, 100%, match=987) putative" /codon_start=1 /protein_id="BAC28767.1" /db_xref="GI:26330057" /translation="MDKICVVGDAVGKTCILISYTTNAPEGEYIPVFDNYSNVMDGKPNVGLMDTAQEDYDRLPISYDQDFLFCFSLVSPASENRAKYPYVRHGPNPPIIVGTRKLDLRDKDIERLKKLPITYPOGLMAKEIGAVKYLECSALTORGLTVEDEAIRAVICPPPVKKRRKCLLL" 2252, .2257 /note="putative" 2269
polyA_site	polyA_site 2269
BASE COUNT	555 a 569 c 502 g 643 t
ORIGIN	
Query Match	67.3%; Score 851.6; DB 11; Length 2269;
Best Local Similarity	85.7%; Pred. No. 1.1e-102;
Matches 1081; Conservative	0; Mismatches 129; Indels 52; Gaps 10;
Db	50 ATGATTCACAGATTTTATTAAGTCATACATGCAAAACACTACTGCTAAGTACAGCAAAA 109
Db	2269 ATGATTCACAGATTTTATTAAGTCATACATGCAAAACACTACTGCTAAGTACAGCAAAA 2210
Qy	110 GATCATATGTAACAACTCCACAAATTCGTGCACTGCAATGTAACAAAGTTTGTCTACT 169
Db	2209 GATCAATGTAACAACTCCACAAATTCGTGCACTGCAATGTAACAAAGTTTGTCTACT 2150
Qy	170 GGTCGAAGGCCCAACACGTGTTCTTGGCAGAGTAGTGTGTACAGAGCGCTTGG 229
Db	2149 GGTTGAAGGCCCAACACGTGTTCTTGGCAGAGTAGTGTGTACAGAGCGCTTGG 2090
Qy	230 CACTAGCGCTTGACAGAACCTCAGACGCCAAG-----TACC 268
Db	2089 CACTAGCAGTTGACAGAACCTCAGACGCCAAGAACATCTTAGCAGAGCAGCATGACA 2030
Qy	269 GGAAGCATGTGTCGGCTGGGTGAGGTGAGAGGGCGGCATCAATCAGATCAGAGTGT 328
Db	2029 GGAAGCATGTGTCGGCTGGGTGAGGTGAGAGGGCGGCATCAATCAGATCAGAGTGT 1973
Qy	329 TGGTACTCTGGCAGACAGATGATGTTTCAATATATCAATATGTTTAAAACTGTAAG 388
Db	1972 TGGTACTCTGGTAAAGGTTTCAATATATCAATATGTTTAAAACTGTAAG 1913
Qy	389 CCGCAGACAGCTGATTTCTACACCCAGTTACTAGAAAACGAAGGAAAGCACTAGT-CAGCT 447
Db	1912 CTGCAACACATGATTTCTACACCCAGTTACTAGAAAACGAAGGAAAGCACTAGTCT 1853
Qy	448 GAGTAAAGAGAGTGAAGACAGACAGCAGCTTCTACTATCTACCAAAAAATCTCCGAA- 506
Db	1852 GAGTAAAGTGAAGCGAAGACAGACAGCAGCTTCTACTAGCTACCAAAAAACAAAAA 1793
Qy	507 ----TGATATATAGAAAGATCTTATATAGTACAGTACAGATATTTGCTGTAAAGAG 562
Db	1792 TCTCTGCTTATCAGAAAGATCTTATATAGTACAGTACAGATATTTGCTGTAAAGAA-GG 1734
Qy	563 GGTCTTAAAGAAAGCACTGTGAAGTTAGCACTGTGAGAGATGCGCAGTTAAATATG 622
Db	1733 GGTCTTAAAGAAAGCACTGTGAAGTTAG-MACTGTGAGAGGTGGCCAGTTTAAATATG 1675

OY		623	ACTAAGCCCCCATCTGCGGGAGGGACACAGGGGGAAGAAGGGGGCCATCAAG-----	671
Db		1674	ACTCAGCCCCCATCTGCGGGAGGGACAGCGCACTGAGGGGTGGGGGCCGAAGGCATGCTC	1615
OY		672	-AGAGACACTGATNAGATCGGCCATTGTTCTACTGTGTTTGACAGAAATTAACCGTTAA	730
Db		1614	AAGAGACACTGATNAGATCGGCCATTGTTCTACTGTGTTTGACAGAAATTAACCGTTAA	1555
OY		731	AAAGCTTACCCTGACACTTTTATTCAGTTGAATTAATCCATGTACATGTAGTGTAAA	790
Db		1554	AAAGCTTACCCTGACACTTTTATTCAGTTGAATTAATCCATGTACATGTAGTGTAAA	1495
OY		791	TTAATCTTACTTATATTAATTAAGTCAAAAAATCTGTCGTCTCCCTTGATACGACGCTTC	850
Db		1494	GTAATCTTACTTATATTAATTAAGTCAAAAAATCTGTCGTCTCCCTTGATATGTGTTTT	1433
OY		851	ACACACTCCACCCAGCACACCCACGACTAGAGACAGAAATACTGCTTAGAGGCAACACAG	910
Db		1435	-CACACTCCACCCAGCACACCCACACAACTAGAGACAGAAATACTGCTTAGAGGCAACACAG	1377
OY		911	GAGCCAGAGTCTGTTCAAAGCCTGCGAGAACCCGCTCAGCTGCTGATTTTAAAGAACTCAC	970
Db		1376	GAAACAGAGTCTGTTCAAAGCTGCGCAAAACGCTAGTCGGCTGCTTTTAAAAAACCTCAC	1317
OY		971	TATGAAATCAAAAGACAGAGCTGTTTACACCCATC---GTGACGTACAGTACAAGTTACG	1027
Db		1316	TATGAAATCAAAAGACAGCTGAGCTGTGCACCTCATCTGACGTGACGTACATACAAAGTTACG	1255
Db		1256	TAAAGAACATGGCTGATTAAGTTACAGGTGCGTTACANIGCAAGCTGTCATTAAAGAGAC	1197
OY		1028	TAAAGAACATGGCTGATTAAGTTACAGGTGCGTTACATGAGCAGGCTGCATTAAAGAGAC	1087
Db		1256	TAAAGAACATGGCTGATTAAGTTACAGGTGCGTTACANIGCAAGCTGTCATTAAAGAGAC	1197
OY		1088	TGTCTCTGTGTACACAGGTCTGGGAGCTACAGGGAGGTTGCAACCCCTGACGCCAGAACT	1147
Db		1196	TGTCTCTGTGTACACAGGTCTGGGAGCTGTGAAAGGCTGTGCACCCCTGCTCAGAAACT	1137
OY		1148	GCACTCTTCTTAAAGCAAA--GTCTCTCAACAGCTTAGTGCTTACGTCTTCACGAC	1204
Db		1136	CGCGTCTCTTACGACAGAAAGTCTCTCAAGAGATTAGTGTAGTGTCTTCACGAC	1077
OY		1205	AAGCCAACTTAGTTCACAAAGGATATTTTGGCAATTTCTAATCTGAGCAAGAAATAGGGAT	1264
Db		1076	AAGCCAACTTAGTTCACAAAGGATATTTTGGCAATTTCTAATCTGAGCAAGAAATAGGGAT	1017
OY		1265	TT 1266	
Db		1016	TT 1015	
<hr/>				
	RESULT 2			
	LOCUS AK081613/c	1796 bp	mRNA	linear HFC 05-DEC-2002
	DEFINITION Mus musculus 16 days embryo head cDNA, RIKEN full-length enriched library, clone:C130051H24 product:Ras-related C3 botulinum substrate 1, full insert sequence.			
	ACCESSION AK081613			
	VERSION AK081613.1	GI:26349264		
	KEYWORDS HFC; CAP trapper.			
	SOURCE Mus musculus (house mouse)			
	ORGANISM Mus musculus			
	REFERENCE 1			
	AUTHORS Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,H., Kono,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.			
	JOURNAL Meth. Enzymol. 303, 19-44 (1999)			
	MEDLINE 99279253			
	PUBMED 10349636			
	REFERENCE 2			
	AUTHORS Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,H., Kono,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.			
	TITLE Normalization and subtraction of cap-trapper-selected cDNAs to			

JOURNAL MEDLINE PUBMED REFERENCE AUTHORS	<p>prepare full-length cDNA libraries for rapid discovery of new genes Genome Res. 10 (10), 1617-1630 (2000)</p> <p>20499374</p> <p>11042159</p> <p>3</p> <p>Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Kono, H., Akiyama, J., Nishi, K., Kitsuana, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujisake, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Matakaki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsutera, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kita, A. and Hayashizaki, Y.</p> <p>Riken integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multipipillary sequencer</p> <p>Genome Res. 10 (11), 1757-1771 (2000)</p>
TITLE	
JOURNAL MEDLINE PUBMED REFERENCE AUTHORS	<p>4</p> <p>Kawai, J., Shingawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y., Aizawa, T., Hara, A., Fukunishi, Y., Kono, H., Adachi, J., Fukuda, S., Atakawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamanaoka, I., Saito, T., Okazaki, Y., Gotohori, T., Bono, H., Kasukawa, T., Saito, R., Kodate, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T., Pleischmann, W., Gaasterland, T., Gissi, C., King, B., Kochwa, H., Knehl, P., Lewis, S., Matsuo, Y., Nukaido, I., Pesole, G., Quackenbush, J., Schriml, L.M., Stikaldi, F., Suzuki, R., Tomita, M., Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H., Baldarelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N., Carninci, P., de Bonaldo, M.F., Brownstein, M.J., Bult, C., Fletcher, C., Fujita, M., Gariboldi, M., Gustlincich, S., Hill, D., Hofmann, M., Hume, D.A., Kamliya, M., Lee, N.H., Lyons, P., Marchionni, L., Mashima, J., Mazzarelli, J., Mombaerts, P., Nordone, P., Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H., Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K.F., Suzuki, H., Toyokawa, K., Wang, R.H., Weitz, C., Willeker, C., Wilming, L., Wynshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawaji, H., Kohlschl, S. and Hayashizaki, Y.</p> <p>Functional annotation of a full-length mouse cDNA collection</p> <p>Nature 409 (6821), 685-690 (2001)</p> <p>21085660</p> <p>11217851</p> <p>5</p>
TITLE	
JOURNAL MEDLINE PUBMED REFERENCE AUTHORS	<p>The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.</p> <p>Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs</p> <p>Nature 420, 563-573 (2002)</p> <p>6 (bases 1 to 1796)</p> <p>Adachi, J., Aizawa, K., Akiyama, T., Atakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashizaki, K., Hayatsu, N., Hiramoto, K., Hirooka, T., Hirose, T., Horii, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Kono, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, Y., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohnato, N., Okazaki, Y., Saito, R., Satoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shingawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.</p>
TITLE	
JOURNAL	<p>Direct Submission</p> <p>Submitted (16-APR-2002) Yoshihide Hayashizaki. The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suifu-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)</p> <p>cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.</p>
COMMENT	

Please visit our web site for further details.

URL: <http://genome.gsc.riken.go.jp/>
URL: <http://fantom.gsc.riken.go.jp/>

FEATURES

Source

Location/Qualifiers

1..1796

/organism="Mus musculus"

/mol_type="mRNA"

/strain="C57BL/6J"

/db_xref="FANTOM_DB:C130051H24"

/db_xref="taxon:10090"

/clone="C130051H24"

/issue_type="head"

/clone_lib="Riken full-length enriched mouse cDNA library"

/dev_stage="16 days embryo"

<1..304

/note="unlabeled protein product; RAS-related C3 botulinum

substrate 1 (MCD) MGI:97845, GBLAK01072, evidence: BLASTW,

100%, match=987)

putative"

/protein_id="BAC38272.1"

/db_xref="GI:26349265"

/translation="VRAKMYPEVRHHCNPNTILVGTKLDRDQKDIIEKLEKLLP

ITVPGIAMAKEIGAVKYLECSALTORGKTFDEAIVRAVLCPPVYKRRKCLLL"

1781..1786

/note="putative"

polya_site 1796

/note="putative"

BASE COUNT 470 a 413 c 371 g 542 t

ORIGIN

Query Match 67.2%; Score 850.2; DB 11; Length 1796;

Best Local Similarity 85.6%; Pred. No. 2e-102;

Matches 1080; Conservative 0; Mismatches 128; Indels 53; Gaps 10;

```

QY 52 GATTCAAGATTATTAAGTCATACATGCAAAACATCTGCTAATCTGATAGCAAAAGA 111
    |||||
DB 1796 GATTCAAGATTATTAATTCATATCATGATGCAAAACATCTGCTAATCTGATAGCAAAAGA 1737
    |||||
QY 112 TCATGTAAGAAACACTGCAATTTCTGCAATTTGTAAGAAAGTTTGTCTAGTGG 171
    |||||
DB 1736 TCATGTAAGAAACACTGCAATTTCTGCAATTTGTAAGAAAGTTTGTCTAGTGG 1677
    |||||
QY 172 TCGAAGAGCCCAACACTGTCTGTCAGAGTGTGTTGTCAGAGAGGGGTTTGA 231
    |||||
DB 1676 TTGAAGAGCCCAACACTGTCTGTCAGAGTGTGTTGTCAGAGAGGGGTTTGA 1617
    |||||
QY 232 CTAGCGCTTGACAGAACCTCAGAGACCCAAAG-----TACCAG 270
    |||||
DB 1616 CTAGCGCTTGACAGAACCTCAGAGACCCCAAGAACCTCTAGAGAGGACATGACAGG 1557
    |||||
QY 271 AAGCATGTGTCGGCTGTGAGGTCTAGAGAGGGGCGGATCATCTCATGACAGGTGG 330
    |||||
DB 1556 AAGCGTGTGTCAGAGGTGGGAGGTCTAGA--GGCAGCATTTAGTCATGACAGGTGG 1500
    |||||
QY 331 GTAATCTGCAAGACATGATGTTTCAAGATATCTAAATAGTTTAAAACTGTAAGACC 390
    |||||
DB 1499 GTAATCTGCAAGACATGATGTTTCAAGATATCTAAATAGTTTAAAACTGTAAGACC 1440
    |||||
QY 391 GCAAGCAGTATTTCTACACCCAGTACTAGAAAAAGAGAGGAGCACTAGT-CAGTGA 449
    |||||
DB 1439 GCAAGCAGTATTTCTACACCCAGTACTAGAAAAAGAGAGGAGCACTAGTACTAGTGA 1380
    |||||
QY 450 GTAAGAGAGGTGAAGAAAGCAAGCACTCTCTACTATCTACCAAAAAAATCTCCGA--- 506
    |||||
DB 1379 GTAAGAGAGGTGAAGAAAGCAAGCACTCTCTACTATCTACCAAAAAAATCTCCGA--- 1330
    |||||
QY 507 ---TGCAATATCAAGAAAGATCTTATAGTACAGGTCAACATATTTGCTGTTAAGAGGG 563
    |||||
DB 1319 CTGCGGCTTATCAAGAAAGATCTTATAGTACAGGTCAACATATTTGCTGTTAAGAA-GGG 1261
    |||||
QY 564 GTCTTAAAGAAAGCACTTGTCTAAGATTAGCAACTGTGAGAGTGGCAGTTTAAATATGA 623
    |||||
```

```

DB 1260 GTTCTAAAGAAAGCACTTGTCTAAGTTAG-ACCTGTGAGGGGTGCCAGTTTAAATATGCA 1202
QY 624 CTCAGAGCCCATCTGGGGAGGAGACAGAGGGGAAGGGGGCTCAAG----- 671
DB 1201 CTCAGAGCCCATCTGGGGAGGAGACAGAGTTGAGAGGGTGGGGGCCGAAGGAGATGCTCA 1142
QY 672 AGAGACATGATTAAGATCGGCCATTTGTCTATCTACGTGTTGACAGAAATTAACCGTTAAA 731
DB 1141 AGAGACATGATTAAGATCGGCCATTTGTCTACCTACGTGTTGACAGAAATTAACCGTTAAA 1082
QY 732 AACCTTACCCGTCAGACACTTTTATTCAGTTGAATTTCTCCATGTACAAATGATGTTAAAT 791
DB 1081 AACCTTACCCGTCAGACACTTTTATTCAGTTGAATTTCTCCATGTACAAATGATGTTAAAG 1022
QY 792 TAACTCTACTATATTAATTAAGTCAAAATATCTGTCGTCCTTGTGATGAGCGTCGTTTCA 851
DB 1021 TAACTCTACTATATTAATTAAGTCAAAATATCTGTCGTCCTTGTGATGATGATGTCGTTT-- 964
QY 852 CACACTCCAGCCAGCACACCCAGCAGACAGTGAAGAAATGATGTTGATGAGGCAACAGG 911
DB 963 CACACTCCAGCCAGCACACCCAGCAGACAGTGAAGAAATGATGTTGATGAGGCAACAGG 904
QY 912 AGCCAGAGTTCTGTTCAAAAGCTGCAAGAACCCGTCAGTGTATTTAGAGAACTCACT 971
DB 903 AACCAAGATCTGTTCAAAGTCTGCAAAAGTGTGAGTGTGCTTTTGAAGAACTCACT 844
QY 972 ATGAATCAAGAGCAGAGCTGTTCACCCATC---GTGACGTACAGTACAAAGTTACGT 1028
DB 843 ATGAATCAAGAGCAGAGCTGTTCACCCATCAGTGTGACGTACAGTACAAAGTTACGT 784
QY 1029 AATGAGCATGGGCTGATTAATTAAGTGTGAGTGTGATGAGGAGTGTGATTAAGAGGCT 1088
DB 783 AATGAGCATGGGCTGATTAATTAAGTGTGAGTGTGATGAGGAGTGTGATTAAGAGGCT 724
QY 1089 GTGCTGTGTCACAGGCTGTGAGAGTGTGAGAGGCTGTGACCCCTGAGCCCAAGAGCTG 1148
DB 723 GTGCTGTGTCACAGGCTGTGAGAGTGTGAGAGGCTGTGACCCCTGAGCCCAAGAGCTG 664
QY 1149 CAGCTCTTTTAAGSACAA--GTCTCTCAACAGCTTATGCTTACGTTGTCAGACA 1205
DB 663 CAGCTCTTTTAAGSACAA--GTCTCTCAACAGCTTATGCTTACGTTGTCAGACA 604
QY 1206 ACGCACTTATGTCACAAAGTATTTTGGCAATCTTAATCTGAGCAAGATTAAGGGATT 1265
DB 603 ACGCACTTATGTCACAAAGTATTTTGGCAATCTTAATCTGAGCAAGATTAAGGGATT 544
QY 1266 T 1266
DB 543 T 543

RESULT 3
AK047969/c 2270 bp mRNA linear HTC 05-DEC-2002
LOCUS AK047969 Mus musculus 16 days embryo head cDNA, RIKEN full-length enriched
DEFINITION library, clone: C130025F01 product: RAS-related C3 botulinum
subtrate 1, full insert sequence.
ACCESSION AK047969
VERSION AK047969.1 GI:26339063
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1
AUTHORS Carninci, P. and Hayashizaki, Y.
TITLE High-efficiency full-length cDNA cloning
METH. ENZYMOLOGY 303, 19-44 (1999)
MEDLINE 99279253
PubMed 10349636
REFERENCE 2
AUTHORS Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
```

TITLE Normalization and subtraction of cap-trapper-selected cDNAs to
JOURNAL prepare full-length cDNA libraries for rapid discovery of new genes
MEDLINE Genome Res. 10 (10), 1617-1630 (2000)
PUBMED 20499374
REFERENCE 11042159
AUTHORS

3 Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
Konno, H., Akiyama, J., Nishi, K., Katsunuma, T., Tashiro, H., Itoh, M.,
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A.,
Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
Fujitake, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watanabe, M.,
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J.,
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A., and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system -384-format
sequencing pipeline with 384 multicapillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)
JOURNAL 20530913
MEDLINE
PUBMED 11076861
REFERENCE
AUTHORS

4 Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y.,
Arakawa, T., Hara, A., Fukunishi, Y., Konno, H., Adachi, J., Fukuda, S.,
Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamataka, I.,
Saito, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Saito, R.,
Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T.,
Fleischmann, W., Gaasterland, T., Gissi, C., King, B., Kochiwa, H.,
Kuehl, P., Lewis, S., Matsuo, Y., Nikaide, I., Pesole, G.,
Quekquenbush, J., Schriml, L. M., Staubli, F., Suzuki, R., Tomita, M.,
Wagner, L., Mashio, T., Sakai, K., Okido, T., Furuno, M., Aono, H.,
Baldarelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N.,
Carninci, P., de Bonaldo, M. F., Brownstein, M. J., Bul, C.,
Flecher, C., Fujita, M., Kamiya, M., Lee, N. H., Lyons, P.,
Hofmann, M., Hume, D. A., Marzella, M., Mazzarello, S., Hill, D.,
Marchionni, L., Mashima, J., Mazarrelli, J., Momberts, P., Nordone, P.,
Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H.,
Sato, K., Schombach, C., Seta, T., Shibata, Y., Storch, K. F., Suzuki, H.,
Toyo-oka, K., Wang, K. H., Welter, C., Whitlaker, C., Wilming, L.,
Wynshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawai, H., Kohlsuk, S.,
and Hayashizaki, Y.
Functional annotation of a full-length mouse cDNA collection
Nature 409 (6821), 685-690 (2001)
JOURNAL 21085660
MEDLINE
PUBMED 11217851
REFERENCE
AUTHORS

TITLE The FANTOM Consortium and the RIKEN Genome Exploration Research
JOURNAL Group Phase I & II Team.
MEDLINE Analysis of the mouse transcriptome based on functional annotation
PUBMED of 60,770 full-length cDNAs
REFERENCE Nature 420, 563-573 (2002)
AUTHORS

5 Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P.,
Fukuda, S., Furuno, M., Hayashizaki, T., Hara, A., Hashizume, M.,
Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoaka, T., Hirozane, T.,
Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T.,
Kato, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Konda, M.,
Koye, S., Kurikawa, C., Matsuyama, T., Miyazaki, A., Murata, M.,
Nakamura, M., Nishi, K., Nomura, K., Numata, R., Ohno, M., Ohsato, N.,
Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N.,
Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,
Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S.,
Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,
Muramatsu, M., and Hayashizaki, Y.
Direct Submission
Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
Kanagawa 230-0045, Japan (E-mail: genome-res@gsr.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222,
Fax: 81-45-503-9216)
cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Division of Experimental Animal Research in Riken contributed to

prepare mouse tissues.
Please visit our web site for further details.
URL: http://genome.gsc.riken.go.jp/
URL: http://fantom.gsc.riken.go.jp/
Location/Qualifiers

FEATURES
source
1..2270
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM,DB:C130025F01"
/db_xref="taxon:10090"
/clone_id="C130025F01"
/issue_type="head"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="16 days embryo"
200..778
/note="unnamed protein product; Ras-related C3 botulinum
substrate 1 (M6D|MGI:97845, GBAK011072, evidence: BLASTN,
100%, match=987)
putative"

CDS
/protein_id="BAC33703.1"
/db_xref="GI:26339064"
/translation="MQAIKCVVVGDSVGRKTLISYTTNAPGEXIPTVDNYSANV
HYDGRPVNGLMDIYEDYDRLRPSFQDVFIFCSISVSPASEFNRKMYEVR
HHCPTMIVLGTGKIDLRDQDKETKEKRTPTTPYQGLAMKEIGAVKYLECSAL
TORGKTYVDEDAIRAVLCPPPVKKRRKRCULL"
2254..2259
/note="putative"
polyA_signal
polyA_site
BASE COUNT 554 a 569 c 503 g 644 t
ORIGIN

Query Match 67.1%; Score 849.6; DB 11; Length 2270;
Best local similarity 85.6%; Pred. No. 2,1e-102;
Matches 1079; Conservative 0; Mismatches 129; Indels 52; Gaps 10;

52 GATTCAGAGATTATTAAGTCATACATGCAAAACATGCTGACGCTTGGCAAAAGA 111
|||||
2265 GATTCAGAGATTATTAATTCATACATGCAAAACATGCTGACGCTTGGCAAAAGA 2210
|||||
112 TCAATGTAAACACATCCACAAATTCGCAATGTCAATGAAAGATTTGTTCTAGTGG 171
|||||
2209 TCAATGTAAACACATCCACAAATTCGCAATGTCAATGAAAGATTTGTTCTAGTGG 2150
|||||
172 TCGAAAGGCCCAACACATGCTGTTCTGCGACGTAGTGTGACGAAGCGCGTTAGCA 231
|||||
2149 TTGAAAGGCCCAACACATGCTGTTCTGCGACGTAGTGTGACGAAGCGCGTTAGCA 2090
|||||
232 CTAGGCGCTGACAGAACCCACAGACCCCAAGG-----TACCGG 270
|||||
2085 CTAGGCGCTGACAGAACCCACAGACCCCAAGG-----TACCGG 2030
|||||
271 AAGCATGTGTCCGCGTGGTGTGAGTCTAGAGGGGGCGGCATCATCATGACAGTGTG 330
|||||
2029 AAGCGTGTGTCCACGCGGGCGAGGCTGTAGCA---GGCAGCATGTAGTCACATGACAGTGTG 1973
|||||
331 GTACTCTGGCAAGACAGATGATGTTTCAAGATTTCTAAATAGTTTAAATCTGTAAGCC 390
|||||
1972 GTACTCTGGCAAGACAGATGATGTTTCAAGATTTCTAAATAGTTTAAATCTGTAAGCC 1913
|||||
391 GCAGCAGCATTTTCTACACCCAGTTACTAGAAAGAGGAGGAGACGACAGT-CAAGTGA 449
|||||
1912 GCAGCAGCATTTTCTACACCCAGTTACTAGAAAGAGGAGGAGACGACAGT-CAAGTGA 1853
|||||
450 GTAAAGGAGGTGAAAGACGACGACCTTCTACTACTATCTACCAAAAATCTCCGAA--- 506
|||||
1852 GTAAAGGAGGTGAAAGACGACGACCTTCTACTACTATCTACCAAAAATCTCCGAA--- 1793
|||||
507 --TGCAATTTATCGAAAGATCTTATATGACAGGTGACAGATTTGCTGTTAAAGGGGG 564
|||||
1792 TCTGGTTATTCAGAAAGATCTTATATGACAGGTGACAGATTTGCTGTTAAAGG-GGGG 1734
|||||

QY 565 TCCTAAGAAAAGCACTTGTCTAAGTAAAGCAACTGTGTAGAGATGGCCACTTTAAATATGAC 624
 1733 TTCTAAGAAAAGCACTTGTCTAAGTAAAGTAAAGTGTAGAGATGGCCACTTTAAATATGAC 1675
 QY 625 TCACAGCCCATATCTGGGGAGGAGACAGACAGGAGGGGGGCTCAAG-----A 672
 1674 TCACAGCCCATATCTGGGGAGGAGAGGAGAGGAGGGGGGCGGCAAGGATGCTCAA 1615
 QY 673 GAGACACTGATTAAGATCGCCCATTTTGTATCTACTGTTTACAGAAATTAACGCTTAA 732
 1614 GAGACACTGATTAAGATCGCCCATTTTGTACTACTGTTTACAGAAATTAACGCTTAA 1555
 QY 733 AGCTTTACCCGTGACACTTTTATTCAGTTAAATTTACTTCATGTCATTAAGTATTAAT 792
 1554 AGCTTTACCCGTGACACTTTTATTCAGTTAAATTTACTTCATGTCATTAAGTATTAAGT 1495
 QY 793 AATCTGACTTCAATTAATTAAGTCAAAATATCTGCTGCTCCCTTATGATGCGCGTTCAC 852
 1494 AATCTGACTTCAATTAATTAAGTCAAAATATCTGCTGCTCCCTTATGATGCGCGTTCAC 1437
 QY 853 ACACCTCCACCACACACACCACAGCTAGAGAACAGAAATCTTCTTAAAGGCAACACAG 912
 1436 ACACCTCCACCACACACACCACAGCTAGAGAACAGAAATCTTCTTAAAGGCAACACAG 1377
 QY 913 GCCAGAGTTTGTTCAAAGCCTGCAAGAGCCGCTGACCTGATTTTAAAGAACTCACTA 972
 1376 ACCAGAGTTTGTTCAAAGCCTGCAAGAGCCGCTGACCTGATTTTAAAGAACTCACTA 1317
 QY 973 TGAATATCAAGACAGAGAGCTGTATACACCATTC--GTAGCTACAGTACAAAGTTACGA 1029
 1316 TGAATATCAAGACAGAGAGCTGTATACACCATTC--GTAGCTACAGTACAAAGTTACGA 1257
 QY 1030 ATGACATGGCGTGATTAAGTTACAGTGTGCTTACATGAGGACGCTGATTAAGAGGCTG 1089
 1256 ATGACATGGCGTGATTAAGTTACAGTGTGCTTACATGAGGACGCTGATTAAGAGGCTG 1197
 QY 1090 TGTGTGTCAACAGGCTGGGAGCTAGGAGGAGGCTGACCCCTGAGCCCAAGAGCTGC 1149
 1196 TGTGTGTCAACAGGCTGGGAGCTAGGAGGAGGCTGACCCCTGAGCCCAAGAGCTGC 1137
 QY 1150 AGTCTTCTTAAGCAAAA--GTCTCTCAACAGCTTACGTTACGTTCTCAGACAA 1206
 1136 GGTCTTCTTAAGCAAAA--GTCTCTCAACAGCTTACGTTACGTTCTCAGACAA 1077
 QY 1207 CGCACTTACTTACCAAGTATTTTGGCAATTTCTTAATCTGAGCAGACAAATAGGGGATTT 1266
 1076 CGCACTTACTTACCAAGTATTTTGGCAATTTCTTAATCTGAGCAGACAAATAGGGGATTT 1017

RESULT 4
 AK076023/c 2308 bp mRNA linear HTC 07-DEC-2002
 LOCUS
 DEFINITION Mus musculus 10 days embryo whole body cDNA, RIKEN full-length enriched library, clone:2610100016 product:RAS-related C3 botulinum substrate 1, full insert sequence.
 AK076023
 VERSION AK076023.1 GI:26344957
 KEYWORDS HTC; CAP trapper.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.

REFERENCE
 1 Carninci, P. and Hayashizaki, Y.
 High-efficiency full-length cDNA cloning
 Meth. Enzymol. 303, 19-44 (1999)
 JOURNAL MEDLINE 99279253
 PUBMED 10349636
 2
 TITLE
 AUTHORs
 TITLE
 Normalization and subtraction of cap-trapper-selected cDNAs to

JOURNAL
 MEDLINE 20499374
 PUBMED 11042159
 REFERENCE
 3
 SHIBATA, K., ITOH, M., AIZAWA, K., NAGAOKA, S., SASAKI, N., CARNINCI, P., KONNO, H., AKIYAMA, J., NISHI, K., KITSUNAI, T., TASHIRO, H., ITOH, M., SUMI, N., ISHII, Y., NAKAMURA, S., HASEGAWA, H., KONDO, S., YAMANAKA, I., YAMAMOTO, R., MATSUMOTO, H., SAKAGUCHI, S., IKEGAMI, T., KASHIWAGI, K., FUJIIWAKE, S., INOUE, K., TOGAWA, Y., IZAWA, M., OHARA, E., MATSUKI, M., YONEDA, Y., ISHIKAWA, T., OZAWA, K., TANAKA, T., MATSUURA, S., KAWAI, J., OKAZAKI, Y., MURAMATSU, M., INOUE, Y., KIRA, A. and HAYASHIZAKI, Y.
 RIKEN Integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer
 Genome Res. 10 (11), 1757-1771 (2000)
 JOURNAL MEDLINE 20530913
 PUBMED 11076861
 4
 KAWAI, J., SHINAGAWA, A., SHIBATA, K., YOSHINO, M., ITOH, M., ISHII, Y., ARAKAWA, T., HARA, A., FUKUNISHI, Y., KONNO, H., ADACHI, J., FUKUDA, S., AIZAWA, K., IZAWA, M., NISHI, K., KIYOSAWA, H., KONDO, S., YAMANAKA, I., SAITO, T., OKAZAKI, Y., GOJOBORI, T., BONO, H., KASUKAWA, T., SAITO, R., KADOTA, K., MATSUDA, H., ASHBURNER, M., BATALOV, S., CASAVANT, T., FLEISCHMANN, W., GASTERLAND, T., GISSI, C., KING, B., KOCHIWA, H., KUEHL, P., LEWIS, S., MATSUO, Y., NIKAI, I., PESOLE, G., QUACKENBUSH, J., SCHIRML, L. M., STAUDI, F., SUZUKI, R., TOMITA, M., WAGNER, L., WASHIO, T., SAKEL, K., OKIDO, T., FURUNO, M., AONO, H., BALDARELLI, R., BARSH, G., BLAKE, J., BOFFELLI, D., BOJUNG, N., CARNINCI, P., DE BOMALDO, M. F., BROWNSTEIN, M. J., BULT, C., HOFMANN, M., FUJITA, M., GARIBOLDI, M., GUSTINCICH, S., HILL, D., HOFMANN, M., HUNE, D. A., KAMIYA, M., LEE, N. H., LYONS, P., MARCHIONNI, L., MASHIMA, J., MAZZARELLI, J., MOMBARTS, P., NORDONE, P., KING, B., RINGWALD, M., RODRIGUEZ, I., SAKAMOTO, N., SASAKI, H., SATO, K., SCHONBACH, C., SEYA, T., SHIBATA, Y., STORCH, K. F., SUZUKI, H., TOYO-OKA, K., WANG, K. H., WELTZ, C., WHITAKER, C., WILMING, L., WYNSHAW-BORIS, A., YOSHIDA, K., HASEGAWA, Y., KAWAJI, H., KOHSEKI, S. and HAYASHIZAKI, Y.
 Functional annotation of a full-length mouse cDNA collection
 Nature 409 (6821), 685-690 (2001)
 JOURNAL MEDLINE 21085660
 PUBMED 11217851
 5
 THE FANTOM CONSORTIUM AND THE RIKEN GENOME EXPLORATION RESEARCH GROUP PHASE I & II TEAM.
 Analysis of the mouse transcriptome based on functional annotation
 Nature 420, 563-573 (2002)
 6 (Pages 1 to 2308)
 ADACHI, J., AIZAWA, K., AKAHIRA, S., AKIMURA, T., AONO, H., ARAI, A., ARAKAWA, T., BONO, H., CARNINCI, P., FUKUDA, S., FUKUNISHI, Y., FURUNO, M., HANAGAKI, T., HARA, A., HAYASU, N., HIRAMOTO, K., HIRAKAWA, T., HORI, F., IMOTANI, K., ISHII, Y., ITOH, M., IZAWA, M., KASUKAWA, T., KATO, H., KAWAI, J., KOJIMA, Y., KONNO, H., KOUDE, M., KOYA, S., KURIHARA, C., MATSUYAMA, T., MIYAZAKI, A., NISHI, K., NOMURA, K., NUMAZAKI, R., OHNO, M., OKAZAKI, Y., OKIDO, T., OWA, C., SAITO, H., SAITO, R., SAKAI, C., SAKAI, K., SANO, H., SASAKI, D., SHIBATA, K., SHIBATA, Y., SHINAGAWA, A., SHIRAKI, T., SOGABE, Y., SUZUKI, H., TAGAMI, M., TAGAWA, A., TAKAHASHI, F., TANAKA, T., TEJIMA, Y., TOYA, T., YAMAMURA, T., YAMANAKA, I., YASUNISHI, A., YOSHIDA, K., YOSHINO, M., MURAMATSU, M. and HAYASHIZAKI, Y.
 Direct Submission
 Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gs.c.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)
 cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

COMMENT

MEDLINE 97044477
 PUBMED 8889548
 COMMENT Contact: Soares, MB
 Coordinated Laboratory for Computational Genomics
 University of Iowa
 375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA
 Tel: 319 335 8250
 Fax: 319 335 9565
 Email: bento-soares@uiowa.edu
 The sequence contained an oligo-dT track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. The sequence tag served to verify it as a clone from the normalized osteoblast library cDNA library Preparation: M.B. Soares Lab Clone distribution: clones will be available through Research Genetics (www.resgen.com)
 Seq primer: M13 Forward
 POLYA=yes.

FEATURES
 Source Location/Qualifiers
 1..760
 /organism="Rattus norvegicus"
 /mol_type="mRNA"
 /strain="Sprague-Dawley"
 /db_xref="taxon:10116"
 /clone="UI-R-DRI-cky-h-11-0-UI"
 /dev_stage="adult"
 /lab_host="DH10B (Life Technologies)"
 /clone_lib="UI-R-DRI"
 /note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker. Site 1: Not I; Site 2: Eco RI; The UI-R-DRI library is a normalized Rat osteoblast library (nreo) constructed in pT37 vector according to the procedure described by Bonaldo, Lennon & Soares (Normalization and Subtraction: Two Approaches to Facilitate Gene Discovery. Genome Research 6: 791-806, 1996). The oligonucleotide used to prime first strand synthesis contained the sequence tag AGATATCAA between the Not I cloning site and dT18 stretch. The Rat Osteoblast tissue was provided by Lien & Stein of the University of Massachusetts Medical School.
 TAG_LIB=UI-R-DRI
 TAG_TISSUE=osteoblast
 TAG_SEQ=AGATATCAA"

BASE COUNT 244 a 143 c 173 g 198 t 2 others

ORIGIN
 Query Match 58.2%; Score 736.4; DB 13; Length 760;
 Best Local Similarity 98.7%; Pred. No. 2,4e-87;
 Matches 752; Conservative 0; Mismatches 8; Indels 2; Gaps 1;

QY 34 TTTTCTTTTCTTTTATGATTCAGAGATTATTAGTCATACATGCAAAACATATGCT 93
 |||||||
 Db 1 TTTTCTTTTCTTTTATGATTCAGAGATTATTAGTCATACATGCAAAACATATGCT 60
 |||||||

QY 94 AACGCACTTGAACAAATCAATGATAAACAACCCCACTGCTTCCAGTCAAGTCAATGAA 153
 |||||||
 Db 61 AACGCACTTGAACAAATCAATGATAAACAACCCCACTGCTTCCAGTCAAGTCAATGAA 120
 |||||||

QY 154 AAAAGTTTCTTCTAGTGGTGGAAAGGCCCAACACTGCTTCTTCCAGTCAAGTGGTG 213
 |||||||
 Db 121 AAAAGTTTCTTCTAGTGGTGGAAAGGCCCAACACTGCTTCTTCCAGTCAAGTGGTG 180
 |||||||

QY 214 TACAGAAAGCGGCTTACGACCTAGCGCTTGACAGAACTCAGACAGCCCAAGGTACCGGAG 273
 |||||||
 Db 181 TACAGAAAGCGGCTTACGACCTAGCGCTTGACAGAACTCAGACAGCCCAAGGTACCGGAG 240
 |||||||

QY 274 CATGTGTCGGGCTGGTGGAGTCTAGAGGGGGCGCATCATCATGAGAGAGTTGGTA 333
 |||||||
 Db 241 CATGTGTCGGGCTGGTGGAGTCTAGAGGGGGCGCATCATCATGAGAGAGTTGGTA 300
 |||||||

QY 334 CTCTGGCAAGACACTGATGTTTCAAGAAATATCTAAATAGTTTAAAGTGAAGCCGCA 393
 |||||||

Db 301 CTCTGGCAAGACACTGATGTTTCAAGAAATATCTAAATAGTTTAAAGTGTACAGCCGCA 360
 |||||||

QY 394 GCACGTGATTTCTACACCCAGTTACTAGAAAACGAAGGAGCACTAGTCAGTGTAA 453
 |||||||

Db 361 GCACGTGATTTCTACACCCAGTTACTAGAAAACGAAGGAGCACTAGTCAGTGTAA 420
 |||||||

QY 454 AGAAGGTGAAAACAGAGACGCTTCTACTATCTACCAAAAAAATCTCGAATGCATTA 513
 |||||||

Db 421 AGAAGGTGAAAACAGAGACGCTTCTACTATCTACCAAAAAAATCTCGAATGCATTA 480
 |||||||

QY 514 TCAGAAAATCTATATGATACAGTCAACATATTTGCTGTTAAGAGGGGCTTAAGA 573
 |||||||

Db 481 TCAGAAAATCTATATGATACAGTCAACATATTTGCTGTTAAGAGGGGCTTAAGA 540
 |||||||

QY 574 AAAGCACTTCTTAAGTATGACACTGTGAGAGTGGCCAGTTAAATATGACTCAAGCC 633
 |||||||

Db 541 AAAGCACTTCTTAAGTATGACACTGTGAGAGTGGCCAGTTAAATATGACTCAAGCC 600
 |||||||

QY 634 CATCTGGGAGGAGCAGACAGAGGGGAGGGGCTCAAGAGACACTGATAGATGGCC 693
 |||||||

Db 601 CATCTGGGAGGAGCAGACAGAGGGGAGGGGCTCAAGAGACACTGATAGATGGCC 660
 |||||||

QY 694 ATTGTCATCTACTGTTTGACAGAAATTAACCGTTAAAGCTTTACCGTGACATTT 753
 |||||||

Db 661 ATTGTCATCTACTGTTTGACAGAAATTA--CCGTTAANAGCTTTACCGTGACATTT 718
 |||||||

QY 754 ATTCACTGAATTAATCTCATATCAATGATAGTAAATTAAT 795
 |||||||

Db 719 ATTCACTGAATTAATCTCATATCAATGATAGTAAATTAAT 760
 |||||||

RESULT 6
 BM389059
 LOCUS 687 bp mRNA linear EST 17-JAN-2002
 DEFINITION UI-R-D20-cko-j-07-0-UI.s1 UI-R-D20 Rattus norvegicus cDNA clone
 BM389059
 ACCESSION UI-R-D20-cko-j-07-0-UI 3', mRNA sequence.
 VERSION BM389059.1 GI:18189112
 KEYWORDS EST.
 SOURCE Rattus norvegicus (Norway rat)
 ORGANISM Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Rattus.

REFERENCE 1 (bases 1 to 687)
 Bonaldo,M.F., Lennon,G. and Soares,M.B.
 Normalization and subtraction: two approaches to facilitate gene discovery
 JOURNAL Genome Res. 6 (9), 791-806 (1996)
 MEDLINE 97044477
 PUBMED 8889548

COMMENT Contact: Soares, MB
 Coordinated Laboratory for Computational Genomics
 University of Iowa
 375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA
 Tel: 319 335 8250
 Fax: 319 335 9565
 Email: bento-soares@uiowa.edu
 The sequence contained an oligo-dT track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. The sequence tag served to verify it as a clone from the non-normalized cartilaginous tumor library cDNA library Preparation: M.B. Soares Lab Clone distribution: clones will be available through Research Genetics (www.resgen.com)
 Seq primer: M13 Forward
 POLYA=yes.

FEATURES
 Source Location/Qualifiers
 1..687
 /organism="Rattus norvegicus"
 /mol_type="mRNA"
 /strain="Sprague-Dawley"

```

/db_xref="taxon:10116"
/clone="UI-R-D20-cko-j-07-0-UI"
/dev_stage="adult"
/lab_host="DH10B (Life Technologies)"
/clone_idb="UI-R-D20"
/notes="Vector: pRT3D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; UI-R-D20 is a
non-normalized Rat cartilaginous tumor library (RCT)
constructed in pRT37 PAC vector according to the procedure
described by Bonaldi, Lennon & Soares (Genome Research
genome 6: 791-806, 1996). The oligonucleotide used to
prime first strand synthesis contained the sequence tag
CATCTGTGTA between the Not I cloning site and dT18
stretch. The Rat cartilaginous tumor tissue was provided
by Dr Jeff Stevens at the University of Iowa.
TAG_Lip=UI-R-D20
TAG_FISUB=cartilaginous tumor
TAG_SEQ=CATCTGTGTA"

```

Query Match	54.1%;	Score 684.4;	DB 12;	Length 687;
Best Local Similarity	99.9%;	Pred. No. 1.6e-80;		
Matches 685; Conservative	0;	Mismatches 1;	Indels 0;	Gaps 0;

OY	33	TTTTTTTTTTTTTTTTTTTTTGTGATCAAGATTTATTAATCATACATGCAAAACATACGC	92
Db	2	TTTTTTTTTTTTTTTTTATGATTCAGACATTTATTAAGTCATACATGCAAAACATACGC	61
OY	93	TAACTGCATTAAGCAAAAGATCATATGTAAAAACCTCCAAATTTGCAACTGTCAATTGA	152
Db	62	TAACTGCATTAAGCAAAAGATCAATGTAAAAACCTCCAAATTTGCAACTGTCAATTGA	121
OY	153	AAAAAGTTGTCTAGTGGTGAAGGCCCAACACTGTGCTTGCCGCTAGTTAGGTT	212
Db	122	AAAAAGTTGTCTAGTGGTGAAGGCCCAACACTGTGCTTGCCGCTAGTTAGGTT	181
OY	213	GTACAGAAAGCGGTAGCATAGCGCTTGACAGAACTCAGAGACCCAAAGTACCGAA	272
Db	182	GTACAGAAAGCGGTAGCATAGCGCTTGACAGAACTCAGAGACCCAAAGTACCGGAA	241
OY	273	GCATGTGCGCGGTGGGTGAGGTCCTAAGGGGGGGGCATCAATACATAGATGTGGT	332
Db	242	GCATGTGCGCGGTGGGTGAGGTCCTAAGGGGGGGGCATCAATCAATCAGATGTGGT	301
OY	333	ACTGTGGCAAGACAGTGAATTTTCAGATATCTAAATAGTTTAAAACTGTAAAGCCGC	392
Db	302	ACTGTGGCAAGACAGTGAATTTTCAGATATCTAAATAGTTTAAAACTGTAAAGCCGC	361
OY	393	AGCAGCGATTTCTACACCCAGTTACTRGAAGAAAGGAAGCATTAGTCACTGAGTA	452
Db	362	AGCAGCGATTTCTACACCCAGTTACTRGAAGAAAGGAAGCATTAGTCACTGAGTA	421
OY	453	AAGAAAGGTGAAAAAGAGACGCACTTCTACTATCTACCAAAAAATTTCCGAATGCAT	512
Db	422	AAGAAAGGTGAAAAAGAGACGCACTTCTACTATCTACCAAAAAATTTCCGAATGCAT	481
OY	513	ATCAGAAAGATCTTATAGTACAGGTCAAGACATATTGCTGTTAAGAAAGGGGTCCTAAAG	572
Db	482	ATCAGAAAGATCTTATAGTACAGGTCAAGACATATTGCTGTTAAGAAAGGGGTCCTAAAG	541
OY	573	AAAAAGCACTTCCTAAGTTAGCAACCTGTGAGATGGCCGTTAAATATGACCTCAAGCC	632
Db	542	AAAAAGCACTTCCTAAGTTAGCAACCTGTGAGATGGCCGTTAAATATGACCTCAAGCC	601
OY	633	CCATCTGGGGAGGGACAGACAGGGGGAAGGGGGGCTCAAGAGACACTGATAGATCGGC	692
Db	602	CCATCTGGGGAGGGACAGACAGGGGGAAGGGGGGCTCAAGAGACACTGATAGATCGGC	661
OY	693	CATTGTCTACTACTGTTTGACAGAA 718	
Db	662	CATTGTCTACTACTGTTTGACAGAA 687	

RESULT 7					
LOCUS	BQ200122				
DEFINITION	BQ200122	705 bp	mRNA	linear	EST 02-MAY-2002
ACCESSION	U1-R-EB1-clf-1-07-0-01.s1		U1-R-EB1	Rattus norvegicus	cDNA clone
VERSION	U1-R-EB1-clf-1-07-0-01.3				mRNA sequence.
KEYWORDS	BQ200122				
SOURCE	BQ200122.1	GI:20416587			
ORGANISM	EST.				
	Rattus norvegicus (Norway rat)				
	Rattus norvegicus				
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;				
	Rattus.				
REFERENCE	1 (bases 1 to 705)				
AUTHORS	Bonaldi,M.F., Lennon,G. and Soares,M.B.				
TITLE	Normalization and subtraction: two approaches to facilitate gene				
	discovery				
JOURNAL	Genome Res.	6 (9),	791-806	(1996)	
MEDLINE	9704447				
PUBMED	8889548				
COMMENT	Contact: Soares, MB				

Email: dento-soares@uowa.edu.au
The sequence contained an oligo-dT track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. The sequence tag present in the cDNA between the NotI site and the oligo-dT track served to verify it as a clone from the normalized duodenum library cDNA library Preparation: M.B. Soares Lab Clone distribution: clones will be available through Research Genetics (www.resgen.com)
Seq primer: M13 forward
POLYA=yes.

FEATURES	Location/Qualifiers
source	1. .705

```

/organism="Rattus norvegicus"
/mol_type="mRNA"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-EB1-clf-1-07-0-UI"
/dev_stage="adult"
/lab_host="DH10B (Life Technologies)"
/clone_lib="UI-R-EB1"
/notes="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; UI-R-EB1 is a
non-normalized Rat Duodenum library (RDU) constructed in
pT737 PAC vector according to the procedure described by
Bonaldo, Lennon & Soares (genome Research genome 6:
791-806, 1996). The oligonucleotide used to prime first
strand synthesis contained the sequence tag TGTGCTTCAT
between the Not I cloning site and d18 stretch. The Rat
Duodenum tissue was provided by Tom Freeman of the Sanger
Center.
TAG_lib=UI-R-EB1
TAG_TISSUE=duodenum
TAG_SEQ=TCGCTTCAT"

```

Query Match	54.0%	Score 683.8;	DB 13;	Length 705;
Best Local Similarity	99.6%	Pred. No. 1.9e-80;		
Matches 685; Conservative	0;	Mismatches 3;	Indels 0;	Gaps 0;

Qy	33	TTTTTTTTTTTTTTATGATTCAGAGATTATTAAAGTATATACATCGAAAACATCTGC	92
Db	2	TTTTTTTTTTTTTTTATATGATTCAGAGATTATTAAAGTATATACATCGAAAACATCTGC	61
Qy	93	TAACTGCATTAGCAAAAGATCAATGTAAAAACACTCCACATTTCTGCAACTGTCAATTGA	152


```

Db      62  TAACTCATTAGGAAAAAGATCAATGTAACAACTCCACATTTCTGCAACTGTCAATGTA 121
QY      153  AAAAATTTGTTCTAGTGTGCGAAAGGCCCAACACTGTGTTCTGCGAGTAGTATAGTT 212
Db      122  AAAAATTTGTTCTAGTGTGCGAAAGGCCCAACACTGTGTTCTGCGAGTAGTATAGTT 181
QY      213  GTACAGACGGCGGTAGACACTAGCGCTTGACAGACCTCAGACCCAAAGGTAACGGAA 272
Db      182  GTACAGACGGCGGTAGACACTAGCGCTTGACAGACCTCAGACCCAAAGGTAACGGAA 241
QY      273  GCATGTGTCGCGGTGGTGGTGTAGAGGGGGCGGCATCAATCATCATGACATGTTGGT 332
Db      242  GCATGTGTCGCGGTGGTGGTGTAGAGGGGGCGGCATCAATCATCATGACATGTTGGT 301
QY      333  ACTCTGGCAAGACAGTAGTATGTTTCAAGATATCTAAATAGTTTAAAACTGTAAGCGCG 392
Db      302  ACTCTGGCAAGACAGTAGTATGTTTCAAGATATCTAAATAGTTTAAAACTGTAAGCGCG 361
QY      393  AGCAGCTGATTTTCTACACCCAGTTACTAGAAAAGGAGGAGCAGCTAGTACCTGAGTA 452
Db      362  AGCAGCTGATTTTCTACACCCAGTTACTAGAAAAGGAGGAGCAGCTAGTACCTGAGTA 421
QY      453  AAGGAAGGTGAAAAAGGAAGGACACTTCTACTATCTACAAAAAAATCTCGAATGCAAT 512
Db      422  AAGGAAGGTGAAAAAGGAAGGACACTTCTACTATCTACAAAAAAATCTCGAATGCAAT 481
QY      482  ATCAGAAAGATCTTATATAGTACAGGTGACACATATGCTGCTTAAGAAAGGGGGTCTTAAG 541
Db      513  ATCAGAAAGATCTTATATAGTACAGGTGACACATATGCTGCTTAAGAAAGGGGGTCTTAAG 572
QY      482  ATCAGAAAGATCTTATATAGTACAGGTGACACATATGCTGCTTAAGAAAGGGGGTCTTAAG 541
Db      573  AAAAGACATTTGCTTAATAGTACAGGTGACAGTATGCTGCTTAAGAAAGGGGGTCTTAAG 632
QY      542  AAAAGACATTTGCTTAATAGTACAGGTGACAGTATGCTGCTTAAGAAAGGGGGTCTTAAG 601
Db      633  CCATCTGGGAGGAGGACAGCAGGGGGGCTCAAGAGAGACACTGATTAAGATCGGC 692
QY      602  CCATCTGGGAGGAGGACAGCAGGGGGGCTCAAGAGAGACACTGATTAAGATCGGC 661
Db      693  CATTTGTCATCTACTGTTTGACAGAAAT 720
QY      662  CATTTGTCATCTACTGTTTGACAGAAAT 689
Db      662  CATTTGTCATCTACTGTTTGACAGAAAT 689

RESULT 8
B0780117 640 bp mRNA linear EST 26-JUL-2002
LOCUS      UI-R-FF0-cow-j-10-0-UI.s1 UI-R-FF0 Rattus norvegicus cDNA clone
DEFINITION UI-R-FF0-cow-j-10-0-UI 3', mRNA sequence.
ACCESSION B0780117 GI:21988589
VERSION    B0780117.1
KEYWORDS   EST.
SOURCE     Rattus norvegicus (Norway rat)
ORGANISM   Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sclurognathia; Muridae; Murinae;
            Rattus.
REFERENCE  1 (bases 1 to 640)
            Bonaldo,M.F., Lennon,G. and Soares,M.B.
            Normalization and subtraction: two approaches to facilitate gene
            discovery
JOURNAL    Genome Res. 6 (9), 791-806 (1996)
MEDLINE    97044477
PUBMED     8889548
COMMENT    Contact: Soares, MB
            Coordinated Laboratory for Computational Genomics
            University of Iowa
            375 Newton Road, 4156 MEBR, Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: bento-soares@uiowa.edu
            Tissue Procurement: Jeff Stevens
            cDNA Library preparation: Dr. M. Bento Soares, University of Iowa

```

```

FEATURES
    source
        location/Qualifiers
            1..640
            /organism="Rattus norvegicus"
            /mol_type="mRNA"
            /db_xref="taxon:10116"
            /clone="UI-R-FF0-cow-j-10-0-UI"
            /tissue_type="Mixed tissues"
            /dev_stage="Adult"
            /lab_host="DH10B (Life Technologies) (T1 phage resistant)"
            /note="Vector: pT73-Pac (Pharmacia) with a modified
            polylinker; Site.1: EcoR I; Site.2: Not I; UI-R-FF0 is a
            subtracted cDNA library containing the following tissue(s)
            ): Normal cartilage and SR-JMS Tumor Line . The
            subtraction was made according to Bonaldo, Lennon and
            Soares, Genome Research, 6:791-806, 1996. The
            oligonucleotide used to prime the synthesis of
            first-strand cDNA contains a library tag sequence that is
            located between the Not I site and the (dT)18 tail. The
            sequence tags for these libraries are: CTAATGACG,
            CATCTTGTA,
            TAG-LIB-UI-R-FF0
            TAG-TISSUE-rat SRC-JMS tumor line
            TAG-SHO-CATCTTGTA"
BASE COUNT  208 a 121 c 152 g 159 t
ORIGIN
Query Match      50.2%; Score 635.8; DB 13; Length 640;
Best Local Similarity 99.7%; Pred. No. 3.8e-74;
Matches 637; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY      33  TTTTGTGTTTGTGTTTATGATTCAGAGATTTTAAAGTCATACATGCAAAACATCTGC 92
Db      2  TTTTGTGTTTGTGTTTATGATTCAGAGATTTTAAAGTCATACATGCAAAACATCTGC 61
QY      93  TAACTGCATTAAGCAAAAGATCAATGTAAACAACTCCACAATCTGCAACTGTCAATGTA 152
Db      62  TAACTGCATTAAGCAAAAGATCAATGTAAACAACTCCACAATCTGCAACTGTCAATGTA 121
QY      153  AAAAAGTTGTTCTAGTGTGCGAAAGGCCCAACACTGTGTTCTGCGAGTAGTATAGTT 212
Db      122  AAAAAGTTGTTCTAGTGTGCGAAAGGCCCAACACTGTGTTCTGCGAGTAGTATAGTT 181
QY      213  GTACAGACGGCGGTAGACACTAGCGCTTGACAGACCTCAGACCCAAAGGTAACGGAA 272
Db      182  GTACAGACGGCGGTAGACACTAGCGCTTGACAGACCTCAGACCCAAAGGTAACGGAA 241
QY      273  GCATGTGTCGCGGTGGTGGTGTAGAGGGGGCGGCATCAATCATCATGACATGTTGGT 332
Db      242  GCATGTGTCGCGGTGGTGGTGTAGAGGGGGCGGCATCAATCATCATGACATGTTGGT 301
QY      333  ACTCTGGCAAGACAGTAGTATGTTTCAAGATATCTAAATAGTTTAAAACTGTAAGCGCG 392
Db      302  ACTCTGGCAAGACAGTAGTATGTTTCAAGATATCTAAATAGTTTAAAACTGTAAGCGCG 361
QY      393  AGCAGCTGATTTTCTACACCCAGTTACTAGAAAAGGAGGAGCAGCTAGTACCTGAGTA 452
Db      362  AGCAGCTGATTTTCTACACCCAGTTACTAGAAAAGGAGGAGCAGCTAGTACCTGAGTA 421
QY      453  AAGGAAGGTGAAAAAGGAAGGACACTTCTACTATCTACAAAAAAATCTCGAATGCAAT 512
Db      422  AAGGAAGGTGAAAAAGGAAGGACACTTCTACTATCTACAAAAAAATCTCGAATGCAAT 481
QY      513  ATCAGAAAGATCTTATATAGTACAGGTGACACATATGCTGCTTAAGAAAGGGGGTCTTAAG 572
Db      482  ATCAGAAAGATCTTATATAGTACAGGTGACACATATGCTGCTTAAGAAAGGGGGTCTTAAG 541

```


Oy	573	AAAGACATCTCTAAGTAGCACAAGTGTAGATGGCCAGTTTAAATATGACTCAAGGCC	6332
Dd	542	AAAAGCACTCTCTAAGTAGTGACACTGTGATGATGCCCGATTMAATATGACTCAAGGCC	601
Oy	633	CCATCTGGGAGGAGCACAGACGGGGGAAGGGGGCTTCAG	671
Dd	602	CCATCTGGGAGGAGCACAGACGGGGGAAGGGGGCTTCAG	640
<hr/>			
RESULT 9 CA339407/c CA339407/c			
LOCUS NISC_LY02409.y1 NCI CGAP Pr32 Rattus norvegicus cDNA clone			
DEFINITION IMAGE:5622185'5', mRNA sequence.			
ACCESSION CA339407			
VERSION CA339407.1 GI:24557505			
KEYWORDS EST.			
SOURCE Rattus norvegicus (Norway rat)			
ORGANISM Rattus norvegicus			
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.			
REFERENCE 1 (bases 1 to 644) NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap. National Cancer Institute, Cancer Genome Anatomy Project (CGAP), TITLE Tumor Gene Index			
JOURNAL Unpublished COMMENT Contact: Robert Strausberg, Ph.D. Email: cgapbs@mail.nih.gov cDNA Library Preparation: CDNA Library Arrayed by: The I.M.A.G.E. Consortium/LNLN DNA Sequencing by: National Institutes of Health Intramural Sequencing Center (NISC) Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LNLN at: info@image.lnl.nih.gov Plate: LHAM12447 row: G column: 18 Seq primer: M13RP1 reverse primer (AB1). Location/Oualifiers			
FEATURES source 1..644 /organism="Rattus norvegicus" /mol_type="mRNA" /db_xref="taxon:10116" /clone="IMAGE:5622185" /sex="male" /tissue_type="dorsal prostate" /dev_stage="adult, 14 month" /lab_host="DH10B (T1 phage-resistant)" /clone_lib="NCI_CGAP_Pr32" /note="Organ: prostate; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 2 kb. Constructed by Invitrogen. Note: This is a NCI-CGAP library."			
BASE COUNT 145 a 164 c 129 g 206 t			
ORIGIN			
Query Match 49.9%; Score 631.4; DB 14; Length 644; Best Local Similarity 99.8%; Pred. No. 1.4e-73; Matches 632; Conservative 0; Mismatches 1; Indels 0; Gaps 0;			
Oy	113	CAATGTAAAAACACTCCCAATTTTGCACACTGTCAATTTGAAAAAAGTTGTTCTAGTGT	172
Dd	644	CAATGTAAAAACACTCCCAATTTTGCACACTGTCAATTTGAAAAAAGTTGTTCTAGTGT	585
Oy	173	CGAAGGCCCAACACTGTGTTCTTGCAGCTAGTAGTTGTACAGAAGGCGTTAGCAC	232
Dd	584	CGAAGGCCCAACACTGTGTTCTTGCAGCTAGTAGTTGTGTACAGAAGGCGTTAGCAC	525
Oy	233	TAGCGCTTGACAGAACCTCACAGACCAGAAAGTACCGGAACATGTCTCCGCTGGGTGA	292
Dd	524	TAGCGCTTGACAGAACCTCACAGACCAGAAAGTACCGGAACATGTCTCCGCTGGGTGA	465
Oy	293	GCTTAGAGGGGGCGCATCATCATGACAGTGTGTGTACTCTGCGCAAGACATGTATG	352

[illegible]

```

/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-CW0s-cch-9-03-0-UI"
/dev_stage="ADULT"
/lab_host="DH10B (Life Technologies)"
/clone_lib="UI-R-CW0s"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-CW0s
library is a non-normalized library constructed from the
following rat aorta tissues: embryonic day 19, embryonic
day 21, adult day 1, adult day 12, adult day 75, adult day
200. For a detailed description of the library from which
this clone was derived, please visit our web site at
ratest.eng.uiowa.edu. The subtraction has been previously
described in (Bonaldi, Lennon and Soares, Genome Research
6:791-806, 1996)
TAG_LIB-UI-R-CW0s
TAG_TISSUE=rat aorta pool
TAG_SEQ=CTGTAGATC"
BASE COUNT      210 a      121 c      149 g      158 t
ORIGIN

```

```

Query Match      49.68; Score 627.4; DB 12; Length 638;
Best Local Similarity 99.18; Pred. No. 4.8e-73;
Matches 631; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

```

```

QY 33 TTTTATGATTCAGAGATTTTAAAGTCATACATGCAAAACATCTGC 92
DB 2 TTTTATGATTCAGAGATTTTAAAGTCATACATGCAAAACATCTGC 61
QY 93 TAACTGATTAAGCAAAAGATCAATGTAAACACCTCCACAATCTGCACTGCAATGGA 152
DB 62 TAACTGATTAAGCAAAAGATCAATGTAAACACCTCCACAATCTGCACTGCAATGGA 121
QY 153 AAAAAGTTGTCTAGTGTGCAAAAGGCCCAACACTGTGTCTTGCCAGTGAAGTTAGTT 212
DB 122 AAAAAGTTGTCTAGTGTGCAAAAGGCCCAACACTGTGTCTTGCCAGTGAAGTTAGTT 181
QY 213 GTACAGAGCGCGTTAGCACTAGCGCTTGACAGAACTCAGACCCCAAGGATACCGGAA 272
DB 182 GTACAGAGCGCGTTAGCACTAGCGCTTGACAGAACTCAGACCCCAAGGATACCGGAA 241
QY 273 GCATGTGTCCGCGTGGGTGAGTGTCTAGAGAGGCGCGCATCAATCATGACAGTGTGGT 332
DB 242 GCATGTGTCCGCGTGGGTGAGTGTCTAGAGAGGCGCGCATCAATCATGACAGTGTGGT 301
QY 333 ACTCTGGCAAGACAGTGTATCTAGAAATATCTAAATAGTTTAAAACTGTAAAGCGCG 392
DB 302 ACTCTGGCAAGACAGTGTATCTAGAAATATCTAAATAGTTTAAAACTGTAAAGCGCG 361
QY 393 AGCAGCGATTTTCTACACCCAGTTACTAGAAAGGAGGAGACACTAGTCACTGAGTA 452
DB 362 AGCAGCGATTTTCTACACCCAGTTACTAGAAAGGAGGAGACACTAGTCACTGAGTA 421
QY 453 AAGGAAGTGAAGAAAGCAAGCACTTCTACTATCTACAAAAAATCTCCGAATGCAAT 512
DB 422 AAGGAAGTGAAGAAAGCAAGCACTTCTACTATCTACAAAAAATCTCCGAATGCAAT 481
QY 513 ATCAGAAAAGATCTTATAGTACAGCTCAGACATATGTCTGTAAAGAGGGGCTCTTAAG 572
DB 482 ATCAGAAAAGATCTTATAGTACAGCTCAGACATATGTCTGTAAAGAGGGGCTCTTAAG 541
QY 573 AAAAGCACTTGCTAAGTATACCACTGTGAGGATGGCCAGTTTAAATATGACATCAAGCC 632
DB 542 AAAAGCACTTGCTAAGTATACCACTGTGAGGATGGCCAGTTTAAATATGACATCAAGCC 601
QY 633 CCATCTGGGAGGAGACAGAGGGGAAAGGGGGCTCA 669
DB 602 CCATCTGGGAGGAGACAGAGGGGAAAGGGGGCTCA 638

```

RESULT 11
BE111691

```

LOCUS      BE111691      639 bp      mRNA      linear      EST 13-JUN-2000
DEFINITION UI-R-BJ1-avv-f-10-0-UI.s1 UI-R-BJ1 Rattus norvegicus cDNA clone
ACCESSION  BE111691
VERSION    BE111691.1 GI:8503796
KEYWORDS   EST.
SOURCE     Rattus norvegicus (Norway rat)
ORGANISM   Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE  1 (bases 1 to 639)
AUTHORS   Bonaldi,M.F., Lennon,G. and Soares,M.B.
TITLE     Normalization and subtraction: two approaches to facilitate gene
JOURNAL   Genome Res. 6 (9), 791-806 (1996)
MEDLINE   97044477
PUBMED    8889548
COMMENT   Contact: Soares, MB
            Coordinated Laboratory for Computational Genomics
            University of Iowa
            375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: bento-soares@uiowa.edu
            The sequence contained an oligo-dT track that was present in the
            oligonucleotide that was used to prime the synthesis of first
            strand cDNA and therefore this may represent a bonafide poly A
            tail. The sequence tag present in the cDNA between the NotI site
            and the oligo-dT track served to identify it as a clone from the
            normalized ventricle at 13 dpc library cDNA library preparation:
            M.B. Soares Lab Clone distribution: clones will be available
            through Research Genetics (www.resgen.com)
            Seq primer: M13 forward
            POLYA=yes.

```

FEATURES

```

source      1..639
            /organism="Rattus norvegicus"
            /mol_type="mRNA"
            /strain="Sprague-Dawley"
            /db_xref="taxon:10116"
            /clone="UI-R-BJ1-avv-f-10-0-UI"
            /lab_host="DH10B (Life Technologies)"
            /clone_lib="UI-R-BJ1"
            /note="Vector: pT73D-Pac (Pharmacia) with a modified
            polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-BJ1
            library is a subtracted library derived from the following
            tissues: atrium at 16.5 dpc, ventricle at 16.5 dpc, AV
            canal at 16.5 dpc, atrium at 15 dpc, ventricle at 15 dpc,
            AV canal at 15 dpc, ventricle at 13 dpc, and adult heart.
            For a detailed description of the library from which this
            clone was derived, please visit our web site at
            ratest.eng.uiowa.edu. The subtraction has been previously
            described in (Bonaldi, Lennon and Soares, Genome Research
            6:791-806, 1996)
            TAG_LIB-UI-R-BJ1
            TAG_TISSUE=ventricle at 13 dpc
            TAG_SEQ=CAGCGA"

```

```

BASE COUNT      209 a      121 c      151 g      157 t      1 others
ORIGIN

```

```

Query Match      49.38; Score 624.6; DB 10; Length 639;
Best Local Similarity 99.28; Pred. No. 1.1e-72;
Matches 627; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

```

```

QY 33 TTTTATGATTCAGAGATTTTAAAGTCATACATGCAAAACATCTGC 92
DB 1 TTTTATGATTCAGAGATTTTAAAGTCATACATGCAAAACATCTGC 60
QY 93 TAACTGATTAAGCAAAAGATCAATGTAAACACCTCCACAATCTGCACTGCAATGGA 152
DB 61 TAACTGATTAAGCAAAAGATCAATGTAAACACCTCCACAATCTGCACTGCAATGGA 120

```

```
QY 153 AAAAAAGTTGTTCTAGTGGTGAAGGCCCCAACACTGTGTTCTGCCAGTAGTAGGTT 212
DB 121 AAAAAAGTTGTTCTAGTGGTGAAGGCCCCAACACTGTGTTCTGCCAGTAGTAGGTT 180
QY 213 GTACAGAAAGGGGTTAGACACTAGCCCTTGACAGAACTCACAGCCCAAGAGTAGCGGA 272
DB 181 GTACAGAAAGGGGTTAGACACTAGCCCTTGACAGAACTCACAGCCCAAGAGTAGCGGA 240
QY 273 GCATGTGTCGCCGCTGGGTAGAGTCTAGAGGGGGCGGCATCATCATGACAGTGTGGT 332
DB 241 GCATGTGTCGCCGCTGGGTAGAGTCTAGAGGGGGCGGCATCATCATGACAGTGTGGT 300
QY 333 ACTGTGGCAAGACACTGATGTTTTCAGAAATCTTAAATAGTTTAAAACTGTAAAGCCGC 392
DB 301 ACTGTGGCAAGACACTGATGTTTTCAGAAATCTTAAATAGTTTAAAACTGTAAAGCCGC 360
QY 393 AGCAGCTGATTTCTACCCAGTCTACTAGAAAAACGAAGGAAGCACTAGCTGAGTA 452
DB 361 AGCAGCTGATTTCTACCCAGTCTACTAGAAAAACGAAGGAAGCACTAGCTGAGTA 420
QY 453 AAGGAAGGTGAAGAACAGAGACACTTCTACTATACCAAAAAATCTCCGATGCAAT 512
DB 421 AAGGAAGGTGAAGAACAGAGACACTTCTACTATACCAAAAAATCTCCGATGCAAT 480
QY 513 ATCAGAAAGATCTTATAGTACAGTCTCAGACATATTTCTCTTAAAGAGGGGTCCTAAAG 572
DB 481 ATCAGAAAGATCTTATAGTACAGTCTCAGACATATTTCTCTTAAAGAGGGGTCCTAAAG 540
QY 573 AAAAGCACTTGCTAGTACGAACTGTGAGATGGCCAGTTTAAATATGAGACTCAAGCCG 632
DB 541 AAAAGCACTTGCTAGTACGAACTGTGAGATGGCCAGTTTAAATATGAGACTCAAGCCG 600
QY 633 CCATCTGGGAGGAGCAGCAGGAGGAGGAGG 664
DB 601 CCATCTGGGAGGAGCAGCAGGAGGAGGAGG 632
```

RESULT 12
BO780658 636 bp mRNA linear EST 26-JUL-2002
LOCUS UI-R-FP0-cpc-c-22-0-UI.s1 UI-R-FP0 Rattus norvegicus cDNA clone
DEFINITION UI-R-FP0-cpc-c-22-0-UI 3', mRNA sequence.

ACCESSION BO780658.1 GI:21989130
VERSION BO780658
KEYWORDS EST.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus

REFERENCE 1 (bases 1 to 636)
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene
JOURNAL discovery
PUBMED 8889548
COMMENT Genome Res. 6 (9), 791-806 (1996)

JOURNAL
MEDLINE
PUBMED
COMMENT
Contact: Soares, MB
Coordinated Laboratory for Computational Genomics
University of Iowa
375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: bento-soares@uiowa.edu
Tissue Procurement: Jeff Stevens
CDNA Library preparation: Dr. M. Bento Soares, University of Iowa
CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: DISTRIBUTION: Researchers may obtain clones
from Research Genetics (www.resgen.com).
Seq primer: M13 FORWARD
POLA=yes.

FEATURES
Location/Qualifiers

```
source 1..636
/organism="Rattus norvegicus"
/mol_type="mRNA"
/db_xref="taxon:10116"
/clone="UI-R-FP0-cpc-c-22-0-UI"
/tissue_type="Mixed tissues"
/dev_stage="Adult"
/lab_host="DH10B (Life Technologies) (T1 phage resistant)"
/clone_lib="UI-R-FP0"
/notes="Vector: pUT73-Pac (Pharmacia) with a modified
polylinker; Site 1: EcoR I; Site 2: Not I; UI-R-FP0 is a
subtracted cDNA library containing the following tissue(s)
: Normal cartilage and SR-TWS Tumor Line . The
subtraction was made according to Bonaldo, Lennon and
Soares, Genome Research, 6:791-806, 1996. The
oligonucleotide used to prime the synthesis of
first-strand cDNA contains a library tag sequence that is
located between the Not I site and the (dT)18 tail. The
sequence tags for these libraries are: CTATGAGAG,
CATCTGTGA.
TAG_LIB=UI-R-FP0
TAG_TISSUE=cartilage
TAG_SEQ=CTATGAGAG"
BASE COUNT 210 a 120 c 148 g 158 t
ORIGIN
Query Match 49.3%; Score 623.8; DB 13; Length 636;
Best Local Similarity 98.9%; Pred. No. 1.4e-72;
Matches 628; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
QY 33 TTTTATTTTATTTATGATTTCAAGATTTTATTAAGTATATGCAATATGCAAAATATGTC 92
DB 2 TTTTATTTTATTTTATGATTTCAAGATTTTATTAAGTATATGCAATATGCAAAATATGTC 61
QY 93 TTAATGCAATTTACCAAAAGATCAATGTAAACACCTCCAAATTCGCAATCTCAATGA 152
DB 62 TTAATGCAATTTACCAAAAGATCAATGTAAACACCTCCAAATTCGCAATCTCAATGA 121
QY 153 AAAAAAGTTGTTCTAGTGGTGAAGGCCCCAACACTGTGTTCTGCCAGTAGTAGGTT 212
DB 122 AAAAAAGTTGTTCTAGTGGTGAAGGCCCCAACACTGTGTTCTGCCAGTAGTAGGTT 181
QY 213 GTACAGAAAGGGGTTAGACACTAGCCCTTGACAGAACTCACAGCCCAAGAGTAGCGGA 272
DB 182 GTACAGAAAGGGGTTAGACACTAGCCCTTGACAGAACTCACAGCCCAAGAGTAGCGGA 241
QY 273 GCATGTGTCGCCGCTGGGTAGAGTCTAGAGGGGGCGGCATCATCATGACAGTGTGGT 332
DB 242 GCATGTGTCGCCGCTGGGTAGAGTCTAGAGGGGGCGGCATCATCATGACAGTGTGGT 301
QY 333 ACTGTGGCAAGACACTGATGTTTTCAGAAATCTTAAATAGTTTAAAACTGTAAAGCCGC 392
DB 302 ACTGTGGCAAGACACTGATGTTTTCAGAAATCTTAAATAGTTTAAAACTGTAAAGCCGC 361
QY 393 AGCAGCTGATTTCTACCCAGTCTACTAGAAAAACGAAGGAAGCACTAGCTGAGTA 452
DB 362 AGCAGCTGATTTCTACCCAGTCTACTAGAAAAACGAAGGAAGCACTAGCTGAGTA 421
QY 453 AAGGAAGGTGAAGAACAGAGACACTTCTACTATACCAAAAAATCTCCGATGCAAT 512
DB 422 AAGGAAGGTGAAGAACAGAGACACTTCTACTATACCAAAAAATCTCCGATGCAAT 481
QY 513 ATCAGAAAGATCTTATAGTACAGTCTCAGACATATTTCTCTTAAAGAGGGGTCCTAAAG 572
DB 482 ATCAGAAAGATCTTATAGTACAGTCTCAGACATATTTCTCTTAAAGAGGGGTCCTAAAG 541
QY 573 AAAAGCACTTGCTAGTACGAACTGTGAGATGGCCAGTTTAAATATGAGACTCAAGCCG 632
DB 542 AAAAGCACTTGCTAGTACGAACTGTGAGATGGCCAGTTTAAATATGAGACTCAAGCCG 601
QY 633 CCATCTGGGAGGAGCAGCAGGAGGAGGAGG 667
DB 602 CCATCTGGGAGGAGCAGCAGGAGGAGGAGG 636
```


Query Match	Best Local Similarity	Matches	Conservative	Score	DB	Length	978;
402	85.9%	0;	1.9e-69;	94;	Indels	32;	Gaps 8;
953	767;	0;	1.9e-69;	94;	Indels	32;	Gaps 8;
460	767;	0;	1.9e-69;	94;	Indels	32;	Gaps 8;
893	767;	0;	1.9e-69;	94;	Indels	32;	Gaps 8;
512	767;	0;	1.9e-69;	94;	Indels	32;	Gaps 8;
833	767;	0;	1.9e-69;	94;	Indels	32;	Gaps 8;
572	767;	0;	1.9e-69;	94;	Indels	32;	Gaps 8;
775	767;	0;	1.9e-69;	94;	Indels	32;	Gaps 8;
632	767;	0;	1.9e-69;	94;	Indels	32;	Gaps 8;
715	767;	0;	1.9e-69;	94;	Indels	32;	Gaps 8;
680	767;	0;	1.9e-69;	94;	Indels	32;	Gaps 8;
655	767;	0;	1.9e-69;	94;	Indels	32;	Gaps 8;
740	767;	0;	1.9e-69;	94;	Indels	32;	Gaps 8;
595	767;	0;	1.9e-69;	94;	Indels	32;	Gaps 8;
800	767;	0;	1.9e-69;	94;	Indels	32;	Gaps 8;
535	767;	0;	1.9e-69;	94;	Indels	32;	Gaps 8;
860	767;	0;	1.9e-69;	94;	Indels	32;	Gaps 8;
477	767;	0;	1.9e-69;	94;	Indels	32;	Gaps 8;
920	767;	0;	1.9e-69;	94;	Indels	32;	Gaps 8;
417	767;	0;	1.9e-69;	94;	Indels	32;	Gaps 8;
980	767;	0;	1.9e-69;	94;	Indels	32;	Gaps 8;
357	767;	0;	1.9e-69;	94;	Indels	32;	Gaps 8;
1037	767;	0;	1.9e-69;	94;	Indels	32;	Gaps 8;
297	767;	0;	1.9e-69;	94;	Indels	32;	Gaps 8;
1097	767;	0;	1.9e-69;	94;	Indels	32;	Gaps 8;
237	767;	0;	1.9e-69;	94;	Indels	32;	Gaps 8;
1157	767;	0;	1.9e-69;	94;	Indels	32;	Gaps 8;
177	767;	0;	1.9e-69;	94;	Indels	32;	Gaps 8;
1214	767;	0;	1.9e-69;	94;	Indels	32;	Gaps 8;

Db	117	TAGTTGAGAAAGGTATTTGGCAATTTTAACTGTCGACGAAGATGGGAGATTTT	65
RESULT 15			
LOCUS	CA318577/c		
DEFINITION	CA318577	826 bp	mRNA
ACCESSION	U1-M-FW0-cbr-o-14-0-U1_r1 NIH_BMAP_FW0		linear
VERSION	IMAGE: 6813639 5', mRNA sequence.		
KEYWORDS	CA318577.1	GI:24536701	
SOURCE	EST.		
ORGANISM	Mus musculus (house mouse)		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
TITLE	1 (bases 1 to 826)		
JOURNAL	NIH-MC http://mgc.nci.nih.gov/.		
COMMENT	Unpublished		
	Contact: Robert Strausberg, Ph.D.		
	Email: sgabbs-remail.nih.gov		
	Tissue Procurement: Dr. Jim Lin, University of Iowa		
	cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa		
	cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa		
	DNA Sequencing by: Dr. M. Bento Soares, University of Iowa		
	Clone Distribution: MGC clone distribution information can be		
	found through the I.M.A.G.E. Consortium/BLND at:		
	http://image.lnsl.gov		
	This clone was contributed by the Brain Molecular Anatomy Project		
	(BMAP)		
FEATURES	Seq primer: pyx-5.		
source	Location/Qualifiers		
	1..826		
	/organism="Mus musculus"		
	/mol_type="mRNA"		
	/strain="C57BL/6"		
	/db_xref="taxon:10090"		
	/clone IMAGE: 6813639"		
	/tissue="IMAGE: "whole brain"		
	/dev_stage="embryo 13.5, 14.5, 16.5, 17.5dpc"		
	/lab_host="DH10B (T1 phage resistant)"		
	/clone_lib="NIH_BMAP_FW0"		
	/note="Organ: Brain; Vector: pyx-Asc; Site_1: EcoR I;		
	Site_2: Not I; The library was constructed according		
	Bonaldi, Lennon and Soares, Genome Research, 6:791-806,		
	1996. Denatured RNA was size fractionated on a 1% agarose		
	gel. First strand cDNA synthesis was primed with oligo-dT		
	primer containing a Not I site. Double strand cDNA was		
	size selected according to mRNA size fraction, ligated		
	with EcoR I adaptor, digested with NotI and then cloned		
	directionally into pyx-Asc vector. The library tag		
	sequence located between the Not I site and the polyA tail		
	is AGCAGACAG. This library was created for the University		
	Iowa Brain Anatomy Project (BMAP). 'Gene Discovery in the		
	Developing Mouse Nervous System', supported by National		
	Institute of Mental Health (NIMH), Hsinlin Chin, Ph.D.,		
	program coordinator."		
BASE COUNT	226 a 195 c 162 g 242 t	1	others
ORIGIN			
Query Match	44.8%; Score 567; DB 14; Length 826;		
Best Local Similarity	89.7%; Pct. No. 3.3e-65;		
Matches 682; Conservative 0; Mismatches 56; Indels 22; Gaps 6;			
QY	525 TTATAGTACAGAGTCACATATTCCTGCTTAAAGAAGGGGCTCTTAAAGAAAGCACTTGC	584	
Db	826 TTATAGTACAGAGTCACATATTCCTGCTTAAAGAAGGGGCTT-CCTAAGAAAGCACTTGC	768	
OY	585 TTAAGTGAACACTGTGAGATGGCCAGTTTAAATATATGATCAACGCCCATCTGGGGAG	644	
Db	767 TTAAGTGAAG-AAGCTGTGAGGGTGGCCAGAGTAAATATGATCAACGCCCATCTGGGGAG	709	

```
QY 645 GGNACAGAGGGGAGAGGGGGCTCAAG-----AGAGACACTGTATAGATCGGC 692
    ||||| ||| | ||||| | ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 708 GGACGGCAGTGGAGGGGGGCGGAGAGATGCTCAAGAGACACTGTATAGATCGGC 649
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 693 CATTTGTCATCTACTGTTTGACAGAAATTAACGTTAAAAAGCTTACCGTGACACTTT 752
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 648 CATTTGTCACCTACTGTTTGACAGAAATTAACGTTAAAAAGCTTACCGTGACACTTT 589
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 753 TATTCAGTTGAATTAATCTCAGTACAAATGATGTAATTAATCTCAGTACATATTAGT 812
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 588 TATTCAGTTGAATTAATCTCAGTACAAATGATGTAATTAATCTCAGTACATATTAGT 529
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 813 CAAAAATACGTGTCGTCCCTTTGATGACGTCGTGTTCAACACACTCCACCAGCACACC 872
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 528 CAAAAATACGTGTCGTCCCTTTGATGACGTCGTGTTCAACACACTCCACCAGCACACC 471
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 873 ACGACTAGAGAAAGAAATTAATCTGTTAGAGCAACACAGAGCCAGAGTTCTGTTCAAGC 932
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 470 ACAACTAGAGAAAGAAATTAATCTGTTAGAGCAACACAGAGCCAGAGTTCTGTTCAAGT 411
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 933 CTGCAGAGCCGGTCAAGCTGGTATTTTGAAGAACTCACTATGAATCAAGAGAGAGAGCT 992
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 410 CTGCAGAGCCGGTCAAGCTGGTATTTTGAAGAACTCACTATGAATCAAGAGAGAGCT 351
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 993 GTTACACCCATC---GTGAGTACAGTACAAAGTTAGCATGAGCATGGGCTGATAGT 1049
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 350 GTTACACCTCATCACTGTGACGTACAGTACAAAGTTAGCATGAGCATGGGCTGATAGT 291
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1050 TACAGTGCCTTACATGCGACGCGTGCATTAAAGAGAGCTGTGCTGTCAACAGGCTCG 1109
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 290 TACAGTGCCTTACATGCGACGCGTGCATTAAAGAGAGCTGTGCTGTCAACAGGCTCG 231
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1110 GAGCTACGGAGAGGCTGTGACACCCCTGAGCCGAGAGCTGAGTCTTAAAGACAAA-- 1167
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 230 GAACCTGTGGAAGGCTGTGACACCCCTGAGCCGAGAGCTGAGTCTTAAAGACAAA-- 171
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1168 -GTCTCTCAACAGCTAGTGTGTTACGTTCTCTCAGCAACGCAACTTAGTTCACAAGT 1226
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 170 TCTCTCAAGAGAGTTAGTGTGTTCTCTCAGCAACGCAACTTAGTTCACAAGT 111
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1227 ATTTTGGCAATTTCTTAATCTGAGCAAGAAATAGGGGATTTT 1266
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 110 ATTTTGGCAATTTTAACTGAGCAAGAAATAGGGGATTTT 71
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

Search completed: August 25, 2003, 23:35:31
Job time : 2641.53 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 25, 2003, 21:58:43 ; Search time 92.8689 Seconds
(without alignments)
6016.990 Million cell updates/sec

Title: US-09-717-321A-2

Perfect score: 1266

Sequence: 1 ttttttttttttttttcaa.....gagcaagatagggatttt 1266

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued_Patents_NA:*
1: /cgn2.6/ptodata/1/ina/5A.COMB.seq:*
2: /cgn2.6/ptodata/1/ina/5B.COMB.seq:*
3: /cgn2.6/ptodata/1/ina/6A.COMB.seq:*
4: /cgn2.6/ptodata/1/ina/6B.COMB.seq:*
5: /cgn2.6/ptodata/1/ina/PT05.COMB.seq:*
6: /cgn2.6/ptodata/1/ina/Backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	187.2	14.8	605	3	US-09-385-982-483
2	68	5.4	588	3	US-09-385-982-133
3	54.2	4.3	7218	1	US-08-232-463-14
4	49.8	3.9	5852	1	US-07-867-106-2
5	47.8	3.8	2218	4	US-09-205-258-103
6	46.8	3.7	760	4	US-09-205-258-232
7	45.6	3.6	38844	4	US-09-734-675-3
8	45	3.6	1882	3	US-09-370-253-1
9	44.6	3.5	991	3	US-08-924-747-25
10	44.6	3.5	991	3	US-09-247-373B-25
11	44.6	3.5	991	3	US-09-296-715-25
12	43.2	3.4	615	4	US-09-105-542A-2
13	43	3.4	1098	3	US-09-248-335-35
14	42.2	3.3	1507	4	US-09-453-323-1
15	42.2	3.3	8100	4	US-09-554-337-4
16	42.2	3.3	11517	1	US-07-920-281C-1
17	42.2	3.3	11517	1	US-08-466-277-1
18	42.2	3.3	15538	4	US-09-554-337-1
19	41.8	3.3	1332	2	US-09-057-762-1
20	41.8	3.3	1332	3	US-08-326-119A-1
21	41.6	3.3	358	2	US-08-721-488-9
22	41.6	3.3	1342	4	US-09-489-847-89
23	41.6	3.3	1929	3	US-09-146-950-1
24	41.4	3.3	501	4	US-09-601-198-170
25	41.2	3.3	1641	1	US-08-300-903A-8
26	41.2	3.3	1641	1	US-08-988-197-8
27	41.2	3.3	1898	1	US-08-342-411A-1

C 28	41	3.2	966	1	US-08-514-014-7	Sequence 7, Appl
C 29	41	3.2	966	2	US-08-833-823-7	Sequence 7, Appl
C 30	40.8	3.2	1117	3	US-09-247-373B-33	Sequence 33, Appl
C 31	40.8	3.2	1878	4	US-09-465-558-39	Sequence 39, Appl
C 32	40.4	3.2	1736	3	US-09-182-816-22	Sequence 22, Appl
C 33	40.4	3.2	1736	3	US-09-182-816-22	Sequence 22, Appl
C 34	40.4	3.2	1736	3	US-09-471-528-22	Sequence 22, Appl
C 35	40.4	3.2	1736	3	US-09-471-528-22	Sequence 22, Appl
C 36	40.4	3.2	1736	3	US-09-634-530-22	Sequence 22, Appl
C 37	40.4	3.2	1736	3	US-09-634-530-24	Sequence 24, Appl
C 38	40.2	3.2	289	1	US-08-341-568-3	Sequence 3, Appl
C 39	40.2	3.2	289	2	US-08-911-020-3	Sequence 3, Appl
C 40	40.2	3.2	859	1	US-08-345-756-8	Sequence 8, Appl
C 41	40.2	3.2	859	1	US-08-625-198-8	Sequence 8, Appl
C 42	40.2	3.2	2039	1	US-08-345-756-5	Sequence 5, Appl
C 43	40.2	3.2	2039	1	US-08-625-198-5	Sequence 5, Appl
C 44	40	3.2	2246	4	US-09-363-708-3	Sequence 3, Appl
C 45	40	3.2	2246	4	US-09-083-587-3	Sequence 3, Appl

ALIGNMENTS

```
RESULT 1
US-09-385-982-483
; Sequence 483, Application US/09385982
; Patent No. 6262334
; GENERAL INFORMATION:
; APPLICANT: ENDEGE, WILSON O., ET AL.
; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
; FILE REFERENCE: CCNDA-260XX
; CURRENT APPLICATION NUMBER: US/09/385, 982
; EARLIER FILING DATE: 1999-08-30
; EARLIER APPLICATION NUMBER: 09/328,111
; EARLIER FILING DATE: 1999-06-08
; EARLIER APPLICATION NUMBER: 60/117,393
; EARLIER FILING DATE: 1999-01-27
; EARLIER APPLICATION NUMBER: 60/098,639
; EARLIER FILING DATE: 1998-08-31
; NUMBER OF SEQ ID NOS: 544
; SOFTWARE: FASTSEQ for Windows Version 3.0
; SEQ ID NO 483
; LENGTH: 605
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(605)
; OTHER INFORMATION: n = A,T,C or G
US-09-385-982-483

Query Match      14.8%; Score 187.2; DB 3; Length 605;
Best Local Similarity 65.5%; Pred. No. 3.3e-41;
Matches 348; Conservative 0; Mismatches 158; Indels 25; Gaps 5;

OY 265 TACCGAAGACGTCGCGCGGCGGAGGTCGTAGAGGGGGGATCAATACATGCA 324
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 74  TACAGAGCGGTGTCTCCGCGGCGAGGTAAAGG-GGGTCAGTATGCTCAAGTGCA 132
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 325 GTGTGTGACTCTGCGACAGACAGTATGTTTCAGAA-----TATCTTAA 367
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 133 GTCGTGAGTATCTGCGACAGACAGTATGTTAAGAGTTTCATAGTTTAAGATTATCTTAA 192
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 368 AATAGTTTAAACAGTAAAGCCGACGACGATGTTTACACCCAGTTACTAGAAACG 427
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 193 AATATTATTAACATTAAGCTGACACACATGATTTTACCTAGTTACTAGAAACT 252
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 428 AAGGGAAGACAGTACAGTGTAGTAAAGAAAGCAAGCAAGCACTCTACTATC 487
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 253 AAGGGAAGACCTTATAGCTCTGAATA-----AAGTAAAGTAAAGCACTTACTAAT 307
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 488 TACCAAAAAAATCTCGAATGATTTATGCAAGAAAGATCTTATAGTACAGGTACAGACATAT 547
```

Db	308	CGACAAAAAAACCTTCTTAATGCATTATCAGAAAGCTTTTATTAATTAACAAG -GAGGCATATT	366
Oy	548	GCATGTTAAGAAGGGGGCTTAAGAAAAAGCACTGCTAATTTGCAACTGTGAGATGG	607
Db	367	GCATCAGTCAGAAAGGGTTCTTATTAAGAAAAAGCACTTACTTAAGTTACGCACTAACGAAACAA	428
Oy	608	CCAGTTTAAATATGCACTCAACGCCCATCTGAGGAGGACAGCAGAGGGGCAAGGGGGCT	667
Db	427	CCAGTTTAAAGATTAATTAATGCCCCCATTTGGGAGAACATGCGAGGGTTTAAACANNAAN	488
Oy	668	CAAGAGAGACACTGATTAAGATGGGCATTTGTC -ATCTACTGTTTGGACAGAAATTAACG	728
Db	487	GAAAGCATTAAAGAAAACATTTNCTGTTATATCAACAACCTTTNTTTNTATCTACTGAAAT	546
Oy	727	TTTAAAGCTTTACCGTGACACTTTATTCAGTTGAATTACTGCATGTAC	777
Db	547	TGACAAAATTAACCTTTTAAAGTTTAAACCCGAGCACTTNTTCTCTTGTCC	597

```

RESULT 2
US-09-385-982-133
Sequence 133, Application US/09385982
Patent No. 6262334
GENERAL INFORMATION:
APPLICANT: ENEGE, WILSON O., ET AL.
TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
FILE OF INVENTION: PRODUCTS: II
FILE REFERENCE: CCDNA-260XX
CURRENT APPLICATION NUMBER: US/09/385,982
EARLIER FILING DATE: 1999-08-30
EARLIER APPLICATION NUMBER: 09/328,111
EARLIER FILING DATE: 1999-06-08
EARLIER APPLICATION NUMBER: 60/117,393
EARLIER FILING DATE: 1999-01-27
EARLIER APPLICATION NUMBER: 60/099,639
EARLIER FILING DATE: 1998-08-31
NUMBER OF SEQ. ID NOS: 544
SOFTWARE: FASTSEQ for Windows Version 3.0
SEQ ID NO 133
LENGTH: 588
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)..(588)
OTHER INFORMATION: n = A,T,C or G
US-09-385-982-133

```

Query Match	5.4%	Score 68	DB 3	Length 588
Best Local Similarity	56.1%	Pred. No. 5.4e-09		
Matches 157	Conservative	0	Mismatches 113	Indels 10
			Gaps	4
QY	361	TATCTAAATAGTTTAAAAACTGTAAGCCGACGACGATGTTTACACCCAGTTACTA	420	
Db	189	TATNNNAATATNNNNNAANNATTAAACCTGACANANTTGATTTTAAACCACTTACTA	248	
QY	421	GAAGACGAAGGAAGACACTAGTACGCTAGTAAAGAAAGTGGAACACGGAACGACTTC	480	
Db	249	GAAGACCTAAGGAAGACACTNATTACT-----CTGAATNANTNACATGGAAACCTTT	302	
QY	481	TACTATCTACCAAAAAAATCTCCGAATTCGATTTATCGAAGATCTTATAGTACAGGTACG	540	
Db	303	TACTATCTTCACAAAAAACCTTC-TCTGCANTATNNNAAAGTTTATNATACA--ANGNG	360	
QY	541	ACATATTCCTGTTAGAGAAGGGGCTCTAAAGAAAAAGCACTGCTTAAGTTAGACACTGNG	600	
Db	361	GNNATCTCTCATCATANNNGGGTCTATTANANA--CCCGCTAANTNTGCGACTTAC	418	
QY	601	AGGATGGCCAGTTTAAATATGAGACTCAACGCCCCCATCTGG	640	
Db	419	AGAACANCCAGCNTANANATANTTTCATGGCCACTTTGGG	458	

```

1      RESULT 3
2      US-08-232-463-14/c
3      Sequence 14, Application US/08232463
4      Patent No. 5670367
5      GENERAL INFORMATION:
6      APPLICANT: DORNER, F.
7      APPLICANT: SCHEIFLINGER, F.
8      APPLICANT: FALKNER, F. G.
9      TITLE OF INVENTION: RECOMBINANT FOURLPOX VIRUS
10     NUMBER OF SEQUENCES: 52
11     CORRESPONDENCE ADDRESS:
12     ADDRESSEE: Foley & Lardner
13     STREET: 1800 Diagonal Road, Suite 500
14     CITY: Alexandria
15     STATE: VA
16     COUNTRY: USA
17     ZIP: 22313-0299
18     COMPUTER READABLE FORM:
19     MEDIUM TYPE: Floppy disk
20     COMPUTER: IBM PC compatible
21     OPERATING SYSTEM: PC-DOS/MS-DOS
22     SOFTWARE: PatentIn Release #1.0, Version #1.25
23     CURRENT APPLICATION DATA:
24     APPLICATION NUMBER: US/08/232,463
25     FILING DATE:
26     CLASSIFICATION: 435
27     PRIOR APPLICATION DATA:
28     APPLICATION NUMBER: US/07/935,313
29     FILING DATE:
30     APPLICATION NUMBER: EP 91 114 300.6
31     FILING DATE: 26-AUG-1991
32     ATTORNEY/AGENT INFORMATION:
33     NAME: BENT, Stephen A.
34     REGISTRATION NUMBER: 29,768
35     REFERENCE/DOCKET NUMBER: 30472/114 IMMU
36     TELECOMMUNICATION INFORMATION:
37     TELEPHONE: (703)836-9300
38     TELEFAX: (703)683-4109
39     TELEX: 899149
40     INFORMATION FOR SEQ ID NO: 14:
41     SEQUENCE CHARACTERISTICS:
42     LENGTH: 7218 base pairs
43     TYPE: nucleic acid
44     STRANDEDNESS: single
45     TOPOLOGY: linear
46     IMMEDIATE SOURCE:
47     CLONE: PTZgpt-Flis
48     US-08-232-463-14

```

[illegible]

[illegible]

RESULT 4

```

US-07-867-106-2
: Sequence 2, Application US/07867106
: Patent No. 5389526
:
: GENERAL INFORMATION:
: APPLICANT: Slade, Martin B
: APPLICANT: Chang, Andy C M
: APPLICANT: Williams, Keith L
: TITLE OF INVENTION: Improved Plasmid Vectors for Cellular
: TITLE OF INVENTION: Slime Moulds of the Genus Dictyostelium
: NUMBER OF SEQUENCES: 19
:
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 5389526rfs
: STREET: One Liberty Place 46th Floor
: CITY: Philadelphia
: STATE: PA
: COUNTRY: USA
: ZIP: 19103
:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentln Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/07/867,106
: FILING DATE: 19920625
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: AU PJ 7187
: APPLICATION NUMBER: PCT/AU90/00530
: FILING DATE: 02-NOV-1989
: ATTORNEY/AGENT INFORMATION:
: NAME: Feeney, Joanne Longo
: REGISTRATION NUMBER: 35,134
: REFERENCE/DOCKET NUMBER: RICE-0002
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 215-568-3100
: TELEFAX: 215-568-3439
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 5852 base pairs
: TYPE: NUCLEIC ACID
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: ANTI-SENSE: NO
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 2378..5038
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 2378..5038
: US-07-867-106-2

```

Query Match	3.98;	Score 49.8;	DB 1;	Length 5852;
Best Local Similarity	57.38;	Pred. NO. 0.0012;		
Matches 90;	Conservative 0;	Mismatches 67;	Indels 0;	Gaps 0;

OY 1 TTTTCTTTTTTTTCAGSTCCAAAGACATTTTTTTTTTTTAGCATTCGAG 60
||||| ||||| ||||| | ||||| ||||| ||||| |||||
Db 5611 TTTTCTTTTTTTTAAATTTTAAAAATTTTTTTTTTTTTTTTAATTAAAAA 5670
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 61 ATTATTAAGTCATACGAAAACATTCGTCTACCGCATACCAAAAGTCAATGTA 120
||||| | ||| ||| ||| ||| ||| ||| ||| |||

Dd 5671 TTTTATTTTATTATTTTAATAAAATTAACAAATATTAAGTAAAAAACCAACAACAAT 5730

Qy 121 AAACACTGCACAATTCTGCAACTGTCAATTGA AAAA 157
|| | | || | ||| | ||| | |||
Db 5731 AACATATTATAAAAAATTCACAATTAAACAATAATTAAA 5767

RESULT 5

US-09/205-258-103/c
Sequence 103, Application US/09205258
Patent No. 6525174
GENERAL INFORMATION:
APPLICANT: Young et al.
TITLE OF INVENTION: 207 Human Secreted Proteins
FILE REFERENCE: P2007P1
CURRENT APPLICATION NUMBER: US/09/205,258
CURRENT FILING DATE: 1998-12-04
EARLIER APPLICATION NUMBER: PCT/US98/11422
EARLIER FILING DATE: 1998-06-04
EARLIER APPLICATION NUMBER: 60/048, 885
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/049, 375
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048, 881
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048, 880
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048, 896
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/049, 020
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048, 876
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048, 895
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048, 884
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048, 894
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048, 971
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048, 964
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048, 882
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048, 899
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048, 893
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048, 900
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048, 901
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048, 892
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048, 915
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/049, 019
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048, 970
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048, 972
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048, 916
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/049, 373
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048, 875
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/049, 374
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048, 917
EARLIER FILING DATE: 1997-06-06

EARLIER APPLICATION NUMBER:	60/048, 876
EARLIER FILING DATE:	1997-06-06
EARLIER APPLICATION NUMBER:	60/048, 8935
EARLIER FILING DATE:	1997-06-06
EARLIER APPLICATION NUMBER:	60/048, 884
EARLIER FILING DATE:	1997-06-06
EARLIER APPLICATION NUMBER:	60/048, 894
EARLIER FILING DATE:	1997-06-06
EARLIER APPLICATION NUMBER:	60/048, 9711
EARLIER FILING DATE:	1997-06-06
EARLIER APPLICATION NUMBER:	60/048, 966
EARLIER FILING DATE:	1997-06-06
EARLIER APPLICATION NUMBER:	60/048, 882
EARLIER FILING DATE:	1997-06-06
EARLIER APPLICATION NUMBER:	60/048, 899
EARLIER FILING DATE:	1997-06-06
EARLIER APPLICATION NUMBER:	60/048, 892
EARLIER FILING DATE:	1997-06-06
EARLIER APPLICATION NUMBER:	60/048, 915
EARLIER FILING DATE:	1997-06-06
EARLIER APPLICATION NUMBER:	60/049, 015
EARLIER FILING DATE:	1997-06-06
EARLIER APPLICATION NUMBER:	60/048, 970
EARLIER FILING DATE:	1997-06-06
EARLIER APPLICATION NUMBER:	60/048, 977
EARLIER FILING DATE:	1997-06-06
EARLIER APPLICATION NUMBER:	60/048, 918
EARLIER FILING DATE:	1997-06-06
EARLIER APPLICATION NUMBER:	60/049, 372
EARLIER FILING DATE:	1997-06-06
EARLIER APPLICATION NUMBER:	60/048, 875
EARLIER FILING DATE:	1997-06-06
EARLIER APPLICATION NUMBER:	60/049, 374
EARLIER FILING DATE:	1997-06-06
EARLIER APPLICATION NUMBER:	60/048, 883
EARLIER FILING DATE:	1997-06-06
EARLIER APPLICATION NUMBER:	60/048, 897
EARLIER FILING DATE:	1997-06-06
EARLIER APPLICATION NUMBER:	60/048, 894
EARLIER FILING DATE:	1997-06-06
EARLIER APPLICATION NUMBER:	60/048, 974
EARLIER FILING DATE:	1997-06-06
EARLIER APPLICATION NUMBER:	60/048, 965
EARLIER FILING DATE:	1997-06-06
EARLIER APPLICATION NUMBER:	60/048, 965
EARLIER FILING DATE:	1997-06-06
EARLIER APPLICATION NUMBER:	60/048, 877
EARLIER FILING DATE:	1997-06-06
EARLIER APPLICATION NUMBER:	60/048, 878
EARLIER FILING DATE:	1997-06-06
EARLIER APPLICATION NUMBER:	60/070, 922
EARLIER FILING DATE:	1997-12-18
EARLIER APPLICATION NUMBER:	60/092, 922
EARLIER FILING DATE:	1998-07-15
EARLIER APPLICATION NUMBER:	60/094, 657
EARLIER FILING DATE:	1998-07-30
NUMBER OF SEQ ID NOS:	1227
SOFTWARE:	PatentIn Ver. 2.0
SEQ ID NO:	232
LENGTH:	760
TYPE:	DNA
ORGANISM:	Homo sapiens
FEATURE:	

```

: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: CDNA
: HYPOTHETICAL: NO
: ANTI-SENSE: NO
: ORIGINAL SOURCE:
: TISSUE TYPE: SOYBEAN

```


Query Match	3.38;	Score 42.2;	DB 4;	Length 15077;
Best Local Similarity	55.98;	Pred. NO. 0.076;		

Search completed: August 26, 2003, 02:33:12
Job time : 94.8689 secs

Search completed: August 26, 2003, 02:33:12
Job time : 94.8689 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleole - nucleole search, using sw model

Run on: August 25, 2003, 22:04:33 ; Search time 309.402 seconds
(without alignments)
9198.941 Million cell updates/sec

Title: US-09-717-321a-2

Perfect score: 1266
Sequence: 1 ttttttttttttttttcaa.....gagcaagaatagggtatttt 1266

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 1517243 seqs, 1124081892 residues

Total number of hits satisfying chosen parameters: 3034486

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA.*

1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq.*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq.*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq.*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq.*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq.*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq.*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq.*
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq.*
10: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq.*
11: /cgn2_6/ptodata/2/pubpna/US09C_NEW_PUB.seq.*
12: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq.*
13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq.*
14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq.*
15: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq.*
16: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq.*
17: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	458.8	36.2	2941	14	US-10-198-846-9874
C 2	447.6	35.4	2051	9	US-09-925-302-255
C 3	430.6	34.0	1644	14	US-10-153-668-48
C 4	430.6	34.0	1646	14	US-10-153-668-46
C 5	420.2	33.2	1232	10	US-09-880-107-1632
C 6	284.2	22.4	2313	14	US-10-198-846-12919
C 7	195.2	15.4	594	14	US-10-066-543-2864
C 8	193.2	15.3	602	10	US-09-878-178-1654
C 9	193.2	15.3	602	13	US-10-046-935-1654
C 10	193.2	15.3	602	14	US-10-146-502-1654
C 11	187.8	14.8	549	14	US-10-066-543-2878
C 12	187.8	14.8	567	14	US-10-066-543-2964
C 13	187.2	14.8	605	11	US-09-871-161-483
C 14	185.4	14.6	477	14	US-10-066-543-3022
C 15	172.2	13.6	432	14	US-09-960-352-7612
C 16	164	13.0	2167	12	US-10-101-510-500

17	163.6	12.9	353	10	US-09-960-352-11432	Sequence 11432, A
C 18	147.4	11.6	409	14	US-10-066-543-13	Sequence 13, Appl
C 19	126.4	10.0	285	9	US-09-294-0938-506	Sequence 506, Appl
C 20	123.2	9.7	416	10	US-09-960-352-8781	Sequence 8781, Ap
C 21	116.4	9.2	3740	11	US-09-764-891-9984	Sequence 9984, Ap
C 22	108.8	8.6	293	14	US-10-066-543-3077	Sequence 3077, Ap
C 23	98.2	7.8	581	10	US-09-988-598-972	Sequence 972, Appl
C 24	95.4	7.5	201	10	US-09-960-352-4357	Sequence 4357, Ap
C 25	91.6	7.2	565	10	US-09-998-598-296	Sequence 296, Appl
C 26	85.6	6.8	492	10	US-09-878-178-1503	Sequence 1503, Ap
C 27	85.6	6.8	492	13	US-10-046-935-1503	Sequence 1503, Ap
C 28	85.6	6.8	492	14	US-10-146-502-1503	Sequence 1503, Ap
C 29	82.2	6.5	398	11	US-09-918-995-34569	Sequence 34569, A
C 30	68	5.4	588	11	US-09-871-161-133	Sequence 133, Appl
C 31	62.2	4.9	426	10	US-09-960-352-7342	Sequence 7342, Ap
C 32	60	4.7	404	10	US-09-960-352-14206	Sequence 14206, A
C 33	57.6	4.5	1651	14	US-10-198-846-13019	Sequence 13019, A
C 34	52.4	4.1	162	10	US-09-920-300A-1549	Sequence 1549, Ap
C 35	52.4	4.1	162	13	US-10-033-528-1549	Sequence 1549, Ap
C 36	52.2	4.1	424	10	US-09-960-352-11218	Sequence 11218, A
C 37	49.4	3.9	469	11	US-09-918-995-13017	Sequence 13017, A
C 38	48.2	3.8	277	10	US-09-960-352-12673	Sequence 12673, A
C 39	48	3.8	640681	10	US-09-790-988-1	Sequence 1, Appl1
C 40	47.8	3.8	2218	14	US-10-023-282-103	Sequence 103, Appl
C 41	46.8	3.7	408	10	US-09-960-352-6263	Sequence 6263, Ap
C 42	46.8	3.7	760	14	US-10-023-282-232	Sequence 232, Appl
C 43	46.4	3.7	15261	11	US-09-764-891-10186	Sequence 10186, A
C 44	45.8	3.6	1008	9	US-09-780-641-1	Sequence 1, Appl1
C 45	45.8	3.6	3770	9	US-09-925-302-346	Sequence 346, Appl

ALIGNMENTS

RESULT 1	US-10-198-846-9874/c	Sequence 9874, Application US/10198846
?	Publication No. US2003009974A1	
?	GENERAL INFORMATION:	
?	APPLICANT: Lillye, James	
?	APPLICANT: Xu, Yongyao	
?	APPLICANT: Wang, Youzhen	
?	APPLICANT: Steilmann, Kathleen	
?	TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS	
?	TITLE OF INVENTION: FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND	
?	TITLE OF INVENTION: THERAPY OF BREAST CANCER	
?	FILE REFERENCE: MRI-049	
?	CURRENT APPLICATION NUMBER: US/10198, 846	
?	CURRENT FILING DATE: 2002-07-18	
?	PRIOR APPLICATION NUMBER: 60/306, 220	
?	PRIOR FILING DATE: 2001-07-18	
?	NUMBER OF SEQ ID NOS: 14084	
?	SOFTWARE: FastSeq for Windows Version 4.0	
?	SEQ ID NO 9874	
?	LENGTH: 2941	
?	TYPE: DNA	
?	ORGANISM: Homo sapiens	
?	US-10-198-846-9874	
QY	1	TTTTTTTTTTTTTTTCAAGTCCAAACATTTTTTTTTTTTTTTTATGATCAAG 60
QY	2481	TTTATATCAAG 2422
QY	61	ATTATTAAGTATCATGCAAAACATCTGTAATGCAATGCAAAAGATCAATGTA 120
QY	2421	ATTATTAAGTATCATGCAAAACATCTGTAATGCAATGCAAAAGATCAATGTA 2362
QY	121	AAACTCCACATTTCTGCACTGTCATTTGAAAAGTTGTCTAGTGTCGAAGGC 180
QY		

Db 2361 AACACTCCACATTTCTGCACTGTCAATTTAAAAATTCGTGTAGTGGCTGAAGGT 2302
QY 181 CCAACACTGTGTCTTCCAGTAGTAGTGGTTGTACAGAAAGCGGTAGCAGTACGC - 238
Db 2301 CCCACCGTGTATTCGTCCAGTAGTGGTTGTACAGAAAGCATGTACAGCAGTACGCAG 2242
QY 239 TTGACGAAGACCTGACAGACCCAAAGG-----TACCGGAGCATG 277
Db 2241 TTTCAGAAACCTTACAGACCCAAAGAAACATATAGGCAAGCCAGCTACAGAGCGGTG 2182
QY 278 TGTCCCGGTGGGTGAGGTGTAGAGGGGGCGGCATCATACATGACAGTGTGTACTCT 337
Db 2181 TGTCCCGGTGGGGAGGTAAAGA-GGGTCAAGTATGTGTCAAGAGACAGTGTGGTATCT 2123
QY 338 GGCAGACAGTAGT-----TTTCAGAAATCTAAATAGTTAAAAA 380
Db 2122 GGCAGACAGTAGTGTAAAGAGGTTTCAATAGTTAAAGAAATATCTAAATATTTAAAAA 2063
QY 381 CTGTAAAGCGCGGACAGGTATTTCTACACCGCATTTACTGAAAGCAAGGAAAGCATCTA 440
Db 2062 CTATTAAGCTGCAACACATATTTTACACCTAGTACTGAAAGCAAGGAAAGCATCT 2003
QY 441 GTACAGTGTAAAGAGGAGTGAAGAACAGAGACGACCTTCTACTATCTACCAAAAAATC 500
Db 2002 ATTAGCTGTGAATA-----AAGTAACATGGAAGAGCACTTTTACTAATCGACAAAAAAC 1948
QY 501 TCCGATGTGATTTACAGAAAGATCTTATAGTACAGGTGACACATATTTCTGTAAAG 560
Db 1947 TTCTAAAGCATTTATCAGAAAGATTTTATTAATACAG-GAGGCATATTTCTACAGCAG 1889
QY 561 GGGGTCTTAAAGAAAGCACTTGTGTAGTAAAGCACTGTAGAGATGGCCAGTTAAATAT 620
Db 1888 GGGTCTTATTAAGAAAGCACTTACTATAGTAAAGCACTTACTAATCGACAAAAAAC 1829
QY 621 GGACTCAACGCCCATCTGGGAGGAGACAGAGGGGAGGCGGCTCAAGAGACACT 680
Db 1828 GAATTAATGCCCAATTTGGGAGGAGCATGGCAGGTAAAGAAAGAAAGCAAGTTAAGA 1769
QY 681 GATTAAGTCCGCCATTTGTATCTACAGTGTGTGACAGAAATTAACCGTTAAAGCTTAC 740
Db 1768 AACATTTCTGATTAATTAACCAACCTTCTTATCATCTACTGTGACAGAAATTAAC 1709
QY 741 CCGTGACACTTTTATTCAGTTGA-----TACTCATGTACAAATTAAGTAAATTAATCT 797
Db 1708 CTTTAAAGATTTTACCGGTGACACTTTCATTCCTTGTACAAATTAAGTAAATCT 1653
QY 798 CTATTCATATTA-GTCAAAATTAAGTGTGTCTCTCTTGTGATGAGCGTGTTCACACAC 856
Db 1652 CCACTTGTATTTGTCAAAATTAAGTGTGTCTCTTGTGATGA-----CACACAC 1603
QY 857 TTCACCCAGCAACCCAGCACTAGAGACAAATTAAGTGTGTAAGGCAACAGAGACCA 916
Db 1602 CCCACCCGCAACCCAGCACTA--AACGAATTTCTCATTAAGAGAAATA-----GCA 1551
QY 917 GAGTCTGTTCAAAGCTGTGAGAAAGCGGTGAGTGTATTTAGAGAACTCACTATGA 976
Db 1550 GTTCTGTTCAAATTTCTGCAAAAGCTGTGAGAAAGCTGTGTAATGAAATCACAAACAT 1491
QY 977 ATCAAGAGCAGAGCTGTACACCCATGTGACGTACAGTACAGTAAAGTTTACGTAATGACA 1036
Db 1490 GATCCAAAGAGCTGAGCTGTACCTCATCTCATTAAGTAAATGATGTATGTGCGGAGACA 1431
QY 1037 TGGGCTGATTAAGTTTACAGTGTGCGTTACATGAGGAGCGTGTCAATTAAGAGGCTGTGCTG 1096
Db 1430 CGTGTCTTACCTACAGTGTGAGTGTCAATGCAAGGCTTCAATTCGGGAGGCTGTGCTGCT 1371
QY 1097 TCAACAGGCTGTG-----GAGCTACGGGAGGCTGTGACAC--CCTAGGCCAAGAACTG 1148
Db 1370 TTACAGCATGTGAGAAATTAAGAGAGAGAGAGTGTGCAACCTCTCAATGCAAGAACTA 1311
QY 1149 CAGCTTTCTTAAGAGCAAGTCTCTCAACAGCTTATGCTTACGTTTCTCAGACACAC- 1207
Db 1310 CCGTCTTCTCAAGAGCAAGGCTTTTCAAAAGTTCATGCTGCGGTCTCTCGGACACACA 1251

QY 1208 --GCAACTAGTTCACAGAGTATTTTGGCAATTTCTTAATCTGAGCAAGATAGCGG 1261
Db 1250 ATCAGAGTGTAGTCAGAGAGGTATTTTGGCAACTTAAATCTGAACAAAGATGGGG 1195

RESULT 2
US-09-925-302-255/c
; Sequence 255, Application US/09925302
; Patent No. US20020044941A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA104
; CURRENT APPLICATION NUMBER: US/09/925,302
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05918
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 896
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 255
; LENGTH: 2051
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (2)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (50)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (68)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (2027)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (2046)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-925-302-255

Query Match 35.48; Score 447.6; DB 9; Length 2051;
Best Local Similarity 68.28; Pred. No. 2.5e-112;
Matches 876; Conservative 0; Mismatches 324; Indels 84; Gaps 15;

QY 33 TTTTATTTTATTTTATGATTCAGAGATTTTATTAAGTATCATACGAAACAAATCTGC 92
Db 2009 TTTTATTTTATTTTATGATTCAGAGATTTTATTAAGTATCATACGAAACAAATATGC 1950
QY 93 TAACTGATTAAGCAAAAGATCAATGTAAACAACTCCACAAATCTGCACTGTCAATTGA 152
Db 1949 TAAATGATTAAGCAAAAGATCAATGTAAACAACTCCACAAATCTGCACTGTCAATTGA 1890
QY 153 AAAAAGTTTGTGTAGTGTGTGGAAGGCCCAACACTGTGTCTTGTCCAGTGTAGTGT 212
Db 1889 AAAAATTTCTGTGTAGTGTGTGGAAGGCTCCACGCTGTATTTCTGCGCAGTGTAGTGT 1830
QY 213 GTACAGAGCGGTGACAGTGTAGCGC--TTGACAGAACTCACAGACCCAAAG----- 264
Db 1829 GTACAGAGCAATGTGACAGCACTAGCACAGTTTACAGAACTCACAGACCCAAAGAACATC 1770
QY 265 -----TACCGGAAGCATGTCTCCGCGGTGAGTGTAGAGGGGGCGGC 309
Db 1769 AATAGCAAAAGCACTACAGAGAGCGGTGTGCGCGGTGGGCGAGGTAAGA-GGGTCACT 1711
QY 310 ATCAATACATGACAGTGTGTGTACTGTGCGACAGACAGTATG----- 352
Db 1710 ATTGCTCAAGTGAAGAGTGTGTGTATGTGCAAGACAGTATGTAAAGAGTTTCAATAGT 1651
QY 353 TTTCAGAAATATCTAAATATAGTTTAAAGAAAGTAAAGCGGAGACAGTGTATTTTACACC 412

Dd	1650	TTAGAAATTATCTAATAATATTTTAAAAAAGCTATTAAGCTGCACACGATATTTTACACT	1591
Qy	413	AGTTACTAGAAAACGAAGGAGACCTAGTCAGCTGAGTAAAGAAAGGTGAAAACAGAA	472
Dd	1590	AGTTACTAGAAAACGAAGGAGACCTATTAGCTGTGATTA-----AAGTAACATGGAA	1536
Qy	473	CGCACTCTACTATCTACCAAAAAAAATCTCGAATGCAATTTTCAGAAAGATCTTATATA	532
Dd	1535	AGCACTTTACTATCTACCAAAAAAAACCTTTTAAGCATTTATCAGAAAAGATTTATATTA	1476
Qy	533	CAGGTACACATATTGCTCGTTAAGAAAGGGGGTCTTAAGAAAAGACCTTGCTAAGTAG	592
Dd	1475	CAAG-GAGGCATATTGCTCAGTACGAAGGGGTTCTATTAAGAAAAGCACTTACTAAGTAG	1417
Qy	553	CACTGTAGAGATGGCCAGTTTAAATATGACTCAACCCCATCTGGGGAGGACAGCA	652
Dd	1416	CGACTAACGAAACAACCGTTTAAAGATTAATATCCCAATTTGGGGAGGACATGGCA	1357
Qy	653	GGGGGAGGGGGGCTCAAGAGACACTGATAGATCGGCCATTTGTCATCTACTGTTG	712
Dd	1356	GGTGTAAAGAGAAAGAAAGCTTAAGAAACATTTCCGATTAATACCACTTTCTTTCA	1297
Qy	713	ACAGAAATTAACCGTTAAAAAGCTTACCCGTGACACTTTATTCAGTTGA--TACT	769
Dd	1296	TCATCTACTGATTTGACAGAAATTAACCTTTTAGAGTTTATACCCGTGACACTTTCAT	1237
Qy	770	CCATGTACATATGTAGTAAATTAATCTCTACTTCATP-TTAGTCAAAATPCTGTCTGC	828
Dd	1236	CTTGTTACAAATGTATGTGA---AATCTCCACTTCGTAATTTTGTCAAAATATCTGCTTGG	1181
Qy	829	TCCTTTGATGACGTCTGTTTCACACACTCCACCCAGCACCCACGACTAGAACAGAA	888
Dd	1180	TCCTTTGATCA-----CACACACCCACCCGGACACACCCACAGCTA--AAGAGAA	1133
Qy	889	TACTTCGTTAAGGACACACAGGAGCCAGAGTTCGTTCAAAGCTGACAGAACCCGTCA	948
Dd	1132	TTCTTCATTAAGAGAAATA-----GCAGTTCTGTTCAAATCTCCGAAAAGCTGGTCA	1079
Qy	949	GCTGTATTATTAGAGAACTCACTPWTGAANTCAAAAGACAGAGACTGTTACACCCATCTGTA	1008
Dd	1078	GAAAACCTGCTATGAATACAAAGAGACTGATCCAAAAGCTGAGCTGTAAGCTACATCC	1019
Qy	1009	CGTACAGTACAAAGTTACGTATAGACATGGGCTGATTAAGTTACAGTGCCTTACATGC	1068
Dd	1018	ATTACAGATCAATGTTATGTGCGGAAACAGCTGCTCACTCACTGGTGAATCAATGGC	959
Qy	1069	AGCTGTCAATTAAGAGGCTGTGCTGCTCAGACAGGCTGG-----GAGCTACGGGAG	1121
Dd	958	AACGCTTCATTCGGGAGGCTTCTGTGCTTACGATCTGAGAACTACATAGAGAGCAG	899
Qy	1122	GGTCTGCACC-CTGAGCCCGAAGACGTGCACTCTCTTAAGACAAAGCTCTCAACAGC	1180
Dd	898	TGTGTGCACCTCTTAACCTGACAGAGCTACCTCTTCTCAAGAGAGAGTCTTTGCAAG	839
Qy	1181	TTAGTGTCTAGTGTCTCAGACACAC--GCAACTTAGTTCACAAAGTATTTGGCAAT	1237
Dd	838	TTCACTGCTCGGTGTTCTCGGCAACCAACATCAGTGTAGTTCAGAGGATATTTGGCAAC	779
Qy	1238	TCTTAATCTGAGCAAGATAGGGG	1261
Dd	778	TCTTAATCTGAGCAAGATAGGGG	755

RESULT 3

US-10-153-668-48/c

: Sequence 48, Application US/10153668

: Publication No. US20030092616a1

: GENERAL INFORMATION:

: APPLICANT: HONDA, Goichi

: APPLICANT: MATSUDA, Akio

: APPLICANT: MORIMATSU, Shuji

: APPLICANT: ISHIZAWA, Kenya

```

? TITLE OF INVENTION: STAT6 Activating Gene
? FILE REFERENCE: 1254-0207P
? CURRENT APPLICATION NUMBER: US/10/153,668
? CURRENT FILING DATE: 2002-05-24
? PRIOR APPLICATION NUMBER: US 60/293,172
? PRIOR FILING DATE: 2001-05-25
? PRIOR APPLICATION NUMBER: US 60/316,031
? PRIOR FILING DATE: 2001-08-31
? PRIOR APPLICATION NUMBER: US 60/328,403
? PRIOR FILING DATE: 2001-10-12
? PRIOR APPLICATION NUMBER: JP 2001-157043
? PRIOR FILING DATE: 2001-05-25
? PRIOR APPLICATION NUMBER: JP 2001-260681
? PRIOR FILING DATE: 2001-08-30
? PRIOR APPLICATION NUMBER: JP 2001-313175
? PRIOR FILING DATE: 2001-10-10
? NUMBER OF SEQ ID NOS: 488
? SOFTWARE: PatentIn Ver. 2.0
? SEQ ID NO 48
? LENGTH: 1644
? TYPE: DNA
? ORGANISM: Homo sapiens
? FEATURE:
? NAME/KEY: CDS
? LOCATION: (109)..(447)
? US-10-153-668-48

```

Query Match	34.0%	Score 430.6	DB 14	Length 1644
Best Local Similarity	67.8%	Pred. No. 1e-107		
Matches	859	Conservative	0	Mismatches 324; Indels 84; Gaps 15;
QY	50	ATGATTCAAGGATTATTATTAAGTCATACATATGCAAAACATACATGCTAACTAGCTATAGCAAAA	109	
DB	1644	ATGATTCAAGGATTATTATTAAGTCATACATGCAAAACATATGCTAACTAGCTATAGCAAAA	158	
QY	110	GATCAATGTAAAAACATCCACAAATTCGTCAACTGTCAATTGAAAAGTTGTCTAGT	169	
DB	1584	GATCAATGTAAAAACATCCACAAATTCGTCAACTGTCAATTGAAAAGTTGTCTAGT	1522	
QY	170	GSTGAAAGGCCCAACACTGTGTTCTTCCACAGTAGTGTGACAGAAAGCGGTAG	229	
DB	1534	GCGTAAAGGTCGCCAGCTGTATTCTCCCAAGTAGTGAAGTTGTACAGAAACATCGTAG	1465	
QY	230	CACATAGCCG--TTGACAGAACCTCCACAAACCCAAAGG-----TA	266	
DB	1464	CACATAGCACAGTTTACACAAACCTCCACAAACCCAAAGAACATCATAGCAAGGAGCTA	1405	
QY	267	CCGGAAGCATGTGTCCCGGTGGGTAGCTTAGAGGGGCGGCATCAATCAGTACAGT	326	
DB	1404	CAGAGGGGCTGTGTCCGCGTGGCGGCGGTAAAGA-GGGTCAAGTATTGGTCAAGTGACAGT	1348	
QY	327	GTTGTACTCTGGGCAAGACAGTANG-----TTTCAGATATCTATAA	369	
DB	1345	GTGGTAATCTGGCAAGACAGTANGTTTAAGAGGTTCATAGTATTAGAAATTAATCTAAA	1286	
QY	370	TAGTTAAAAACTGTAAAGCGCGACAGCAGTGTATCTACACCAGTATCTAGAAAAAGAA	429	
DB	1285	TATTTTAAAACTATAAAGCTGCAACACATGATTTTACCTAGTATCTAGAAAAACATRA	1226	
QY	430	GGAAGCACTAGTCAGTGAAGTAAAGAAAGTGAACAAGAACGCACTTCTACTATCTA	489	
DB	1225	GGAAGCACTATTAGTCTCTAATP-----AAGTAACTGTGAAGAACACTTTTACTAATCG	1172	
QY	490	CCAAAAAATATCCGAATTCATTAACAAAAGATTTTATAGACAGGTACAGACATATTCG	549	
DB	1170	ACAAAAAATCTTCTAATGCATTAATCAAAAGATTTTATTAATACAG-GAAGCATATTCG	1112	
QY	550	TCGTTAAGAAAGGGGGTCTCTAAAGAAAGACACTTGTCTAATAGCAACTGTAGAGTATGCC	609	
DB	1111	TCAGTCAGAAAGGGGTTCTATAAGAAAGACACTTACTAAGTATAGGACATACAGAAACAC	1051	
QY	610	AGTTAAATATGAGACTCAAGCCCATTTGGGAGGAGCACAGAGGGGAGAGGGGGCTCA	669	

Db 1051 AGTTTAAAGATGATTTAAATGCCCAATTTGGGGAGCATGGCAGCTGTAAAGAGAAAGAA 992
QY 670 AGAGAGACACTGATTAAGATGGCCATTTGTCACTACTGTTCACAGAAATTAACCTTA 729
Db 991 AAGCTTAAGAAACATTTCTGATTAATACCAACCTTTCTTTCATCATCTGATTTGA 932
QY 730 AAAAGCTTACCGGTGACACTTTTATTGAGTTGA--TTACTGCATGTCAATGTAGTG 786
Db 931 CAGAAATTAACCTTTAGAGTTTAACTTTTACCGTGAACATTTCTCTGTCAATGTAGTG 872
QY 787 TAAATTAATCTACTACTCATTA-TTAACTCAAAATACGTGTGTCTGCTTATGACGTG 845
Db 871 TA-----AATCTCCACTCTGATTTTGTCAAAATACGTCTTTGTCTTTATCA----- 822
QY 846 GTTTCACACACTCCACCCACACCCACAGCTAGGAAGAAATACCTTCTGTAGAGCAA 905
Db 821 ----CACACACCCACCCGACACCCACAGCTA--AACGAATTTCTTCAATGAGAGAA 768
QY 906 CACAGGAGCCAGATTTCTGTCAAAAGCTGACAGAGCCGTGACGTGATTTTAGAA 965
Db 767 TAG-----CAGTTCTGTCAAAATCTCCGCAAAAGCTGTGAGAAATCTGCTAGAA 714
QY 966 CTCATATGAATCAAAAGACAGAGCTGTACACCCATCTGACGTACAGTACAAAGTAA 1025
Db 713 TCACAAAGACTGATCCAAAGAGCTGAGCTGCTAGCTCCTCATTTACATACATGTA 654
QY 1026 CGTAATGAGCATGGCTGATTAAGTTCAGAGTGGCTTACATGCGCTGTATTAAGAG 1085
Db 653 TGTGGGGAACACGTGCTGCTACCTGCTGATGATGATCAAGCAACCTTCTGCGGAG 594
QY 1086 GCTGTGCTGTACACAGGTCTGG-----GAGCTACGGAGGAGGTCTGACCC-CCGTAG 1137
Db 593 GCTGTCTGCTTTACGATCTGAGAACTACATAGAGAGCAAGTGTGCACTCTCTAC 534
QY 1138 CCCAAGCTGCACTCTTCTTAAGAGCAAAAGCTCTACAGCTTACTGTCTAGCTTC 1197
Db 533 TCGAAGAGTACCGTCTTCTCAAAAGAGAGTCTTTCGAAAGTTCAGTCTGCTGCTTC 474
QY 1198 TCAGCACAAC---GCAACTTATGTCACAGGATTTTGGCAATTTCTTAATCTGAGCAGA 1254
Db 473 TCGGCAACAAAGTACGATGATGATGAGATTTTGGCAACCTTAACTGAAACAGA 414
QY 1255 ATAGGGG 1261
Db 413 ATGGGGG 407

RESULT 4
US-10-153-668-46/c
; Sequence 46, Application US/10153668
; Publication No. US20030092616A1
; GENERAL INFORMATION:
; APPLICANT: HONDA, Goichi
; APPLICANT: MATSUDA, Akio
; APPLICANT: MURAMATSU, Shuji
; APPLICANT: ISHIZAMA, Kenya
; TITLE OF INVENTION: STAT6 Activating Gene
; FILE REFERENCE: 1254-0207P
; CURRENT APPLICATION NUMBER: US/10/153, 668
; CURRENT FILING DATE: 2002-05-24
; PRIOR APPLICATION NUMBER: US 60/293,172
; PRIOR FILING DATE: 2001-05-25
; PRIOR APPLICATION NUMBER: US 60/316,031
; PRIOR FILING DATE: 2001-08-31
; PRIOR APPLICATION NUMBER: US 60/328,403
; PRIOR FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: JP 2001-157043
; PRIOR FILING DATE: 2001-05-25
; PRIOR APPLICATION NUMBER: JP 2001-260681
; PRIOR FILING DATE: 2001-08-30
; PRIOR APPLICATION NUMBER: JP 2001-313175
; PRIOR FILING DATE: 2001-10-10
; NUMBER OF SEQ ID NOS: 488

SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 46
; LENGTH: 1646
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (109)..(273)
US-10-153-668-46
Query Match 34.0%; Score 430.6; DB 14; Length 1646;
Best Local Similarity 67.8%; Pred. No. 1e-107;
Matches 859; Conservative 0; Mismatches 324; Indels 84; Gaps 15;
QY 50 ATGATTCAGAGATTTTAAAGTATCATATGCAAAACATCTGCTAATCTGATAGCAAAA 109
Db 1646 ATGATTCAGAGATTTTAAAGTATCATATGCAAAACATCTGCTAATCTGATAGCAAAA 1587
QY 110 GATCAATGTAAAAACATCTCCAAATTTCTGCACTGTCAATTTGAAAAAGTTTGTCTAGT 169
Db 1586 GATCAATGTAAAAACATCTCCAAATTTCTGCACTGTCAATTTGAAAAAGTTTGTGTAGT 1527
QY 170 GGTCAAGAGCCCAACACTGTGTCTTCCAGTGTAGTGTGTACAGAAAGCGGTAG 229
Db 1526 GGTCAAGAGGTCCCAAGCTGTATTTCTGCAAGTGTAGTGTGTACAGAAACATCTGAG 1467
QY 230 CACTAGCGC--TTGACAGAACTTCACAGACCCAAAG-----TA 266
Db 1466 CACTAGCAGCTTTACAGAACTTCACAGACCCAAAGCAATCAATAGCCAAAGCGACTA 1407
QY 267 CCGGAAGCATGTCCGCGGTGGGTAGAGGTGTAGAGGGGGGGGATCAATCATATGACAGT 326
Db 1406 CAGAGGCGTGTGTCCGCGGTGGGTAGAGGTGTAGAGGGGGGGGATCAATCATATGACAGT 1348
QY 327 GTTGTACTCTGGCAAGACAGTATG-----TTTCAGAAATCTTAAAA 369
Db 1347 GTGGTAAATCTGGCAAGACAGTATGTTAGAAAGTTTCAATGTTAAGATTTCTTAAA 1288
QY 370 TAGTTTAAAAAGTAAAGCCGACAGTGTATTTTACACCCAGTACTAGTAAAGAAAGAA 429
Db 1287 TATTTTAAAAAGTAAAGCCGACAGTGTATTTTACACCCAGTACTAGTAAAGAAAGTAA 1228
QY 430 GGGAGCAGTACGTAGTGTAAAGAGTGTAAAGAGTGTAAAGAGTGTAAAGAGTGTAAAG 489
Db 1227 GGAAGCAGTATTAATGCTGTGAATA-----AAGTACATGGAAGACCTTTACTAATCG 1173
QY 490 CCAAAAAATCTCGAATGATTAATCAGAAAGATCTTATAGTACAGTGTAGACATATGTC 549
Db 1172 ACAAAAAATCTCTAATGATTAATCAGAAAGATTTTAAATTAATCAAG-GAGCATATGTC 1114
QY 550 TCGTTAAGAGGGGGTCTTAAAGAAAGCACTTGTCTAAGTTACCACTGTGAGATGGCC 609
Db 1113 TCAGTCAGAGGGGTCTTAAAGAAAGCACTTGTCTAAGTTACCACTGTGAGATTAACAAAC 1054
QY 610 AGTTTAAATATGAGCTCAAGCCCATCTGGGGAGGAGCAGAGGGGAGGGGGCTCA 669
Db 1053 AGTTTAAAGTGAATTAATGCCCCAATTTGGGAGGAGTGGCAGGTGTAAAGAAAGAA 994
QY 670 AGAGAGCACTGATTAAGATGGCCATTTGTCACTACTGTGTGACAGAAATTAACCGTTA 729
Db 993 AAGCTTAAGAAACATTTCTGATTAATACCAACCTTTCTTCACTACTCATCTCATTTGA 934
QY 730 AAAAGCTTACCGGTGACACTTTTATTGAGTTGA--TTACTGCATGTCAATGTAGTG 786
Db 933 CAGAAATTAACCTTTAGAGTTTAACTTTTACCGTGAACATTTCTCTGTCAATGTAGTG 874
QY 787 TAAATTAATCTACTACTCATTA-TTAACTCAAAATACGTGTGTCTGCTTGTAGAGCGGT 845
Db 871 TA-----AATCTCCACTCTGATTTTGTCAAAATACGTCTTTGTCTTTATCA----- 824
QY 846 GTTTCACACACTCCACCCACACCCACAGCTAGGAAGAAATACCTTCTGTAGAGCAA 905
Db 823 ----CACACACCCACCCGACACCCACAGCTA--AACGAATTTCTTCAATGAGAGAA 770

QY	608	CCAGTTTAAATATGACTACACGCCCATCTGAGGAGGAGACGACAG-----GGG	657
Db	427	CCAGTTTAAAGATGAATTTAAATGCCCAATTTGGGAGGACATGGTAAAGAAAG	488
QY	658	AAAGGGGGGCTCAAGACAGACACTGATTAAGATGGC-----CATTTGTATCTACAG--TTT	711
Db	487	AAAGCTTAAAGAAACATTTCTCGATTAACACACCTTTCTTATCATCATCATGCAATT	546
QY	712	GACAGAAATTAACCGTTAAAGAGCTTACCGTGCACCTTTAATTC	757
Db	547	GACAGAAATTAACCTTTTAAAGATTTTACCCGTGACATTTTCATTTC	592

```

RESULT 10.502-1654
US-10-146-502-1654
Sequence 1654, Application US/10146502
Publication No. US2003006180A1
GENERAL INFORMATION:
APPLICANT: JIANG, YUQU
APPLICANT: HARLOCKER, SUSAN L.
APPLICANT: SECRETIST, HEATHER
APPLICANT: WANG, AIJUN
APPLICANT: SCOLK, JOHN A.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
OF COLON CANCER
FILE REFERENCE: 210121.527C2
CURRENT APPLICATION NUMBER: US/10/146,502
CURRENT FILING DATE: 2002-05-14
NUMBER OF SEQ. ID NOS: 2241
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO 1654
LENGTH: 602
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: 569..578
OTHER INFORMATION: n = A,T,C or G
US-10-146-502-1654

```

Query Match	Score	DB	Length
15.38;	193.2;	DB 14;	Length 602;

```
Matches 371; Conservative 0; Mismatches 115; Indels 40; Gaps 7;
```

QY	265	TACCGGAAGCATGTGTCGCCCGCTGGGGAGAGCTGTACAGGGGGCGGCGATCAATACATATGACA	32
Db	74	TACAGGAGGGGTGTGTGTCGCCGTGGGGAGGTAAAG-GGGTGAGTATTTGGACAGTCCCA	133
QY	325	GTGTGGTACTCTGGCAGACAGTATGTTTCAGAA-----TATCTAA	367
Db	133	GTGCGGTATCTGGCAGACAGTATGTTTAAGAAAGTTCATAGTTTAAGAATATCTAA	192
QY	368	AATATGTTTAAAACTGTAAAGCCGACAGCAGTATTTTACACCCAGTTACTAGAAAAAG	427
Db	193	AATATTTTAAAACTGTAAAGCTGCACACATGATTTTACACCTAGTTACTAGAAAACT	252
QY	428	AAGGGAAGCCTAGTAGAGTAAAGGAAGCTAAAAAGGAACGACCTCTACATATC	487
Db	253	AAGGAAGCCTATTATAGCTCTGAAATA-----AATTAACATGGAAGACACTTTTACTATAT	307
QY	488	TACCAAAAAAATCTCCGAATGCATTTATCGAAGAATCTTATATAGTACAGTCAGACATATT	547
Db	308	CGACAAAAAAACCTTCTATATGCATTTATCGAAGAATTTTATATATACAAAG-GAGGCATATT	366
QY	548	GCTCGTTTAAAGAGGGGGTCTTAAAGAAAAAGCACTTGCTAAGTTAAGCACTGTGAGGATGG	607
Db	367	GCTCAGTCAGAAAGGGGTTCTTATTAAGAAAAAGCACTTACTAAGTTACGCACTTAACAGAACAA	428
QY	608	CCAGTTTAAATATAGCATCAACGCCCATCTGGGAGGAGGACAGCAGCG-----GGG	657
Db	427	CCAGTTTAAAGATTAATTAATGCCCAATTTTGGGAGGACATGGCAGGGTAAAGAAAGG	486

Oy	658	AAAGGGGGGCTCAGAGAGACATGATTAAGATGGC----	CATTGTATCTACAG--TTT	711
Db	487	AAAAGCTTAGAANAACATTTCTCGTAATACCAACCTTCTTTCATCATCTACTGCATTT		548
Oy	712	GACAGAAATTAACCGTTAAAAAGCTTTACCGTCACATTTATTC	757	
Db	547	GACGAGAAATTAACCTTTTAGAGCTTTTACCCCTGACACTTTCATTC	592	

```

RESULT 11
US-10-066-543-2878
; Sequence 2878, Application US/10066543
; Publication No. US20030087818A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Yuqiu
; APPLICANT: Pyle, Ruth A.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Indrias, Carol Yoseph
; APPLICANT: Lodes, Michael J.
; APPLICANT: Secrist, Heather
; APPLICANT: Carter, Darrick
; APPLICANT: Fanger, Gary R.
; APPLICANT: Smith, Carole L.
; APPLICANT: Durham, Margareta
; APPLICANT: Stolk, John A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF COLON CANCER
; FILE REFERENCE: 210121.563
; CURRENT APPLICATION NUMBER: US/10/066.543
; NUMBER OF SEQ ID NOS: 3417
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2878
; LENGTH: 549
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 526
; OTHER INFORMATION: n = A,T,C or G
US-10-066-543-2878

```

Query Match	14.88; Score 187.8; DB 14; Length 549;
-------------	--

Matches	304;	Conservative	0;	Mismatches	87;	Indels	24;	Gaps	4;
---------	------	--------------	----	------------	-----	--------	-----	------	----

OY	265	TACCGGAAGCATGTGTCGCCGCTGAGAGCTTAAGGGGGCGGCATCAATCATGACA	322
Db	74	TACGGGGGGGTGTGTCCGCTGGGGGAGGTAAAG - GGCTCAGTATTTGTCMAAGTACA	132
OY	325	GTCGTGTACTCTCGGAAGACAGTGAATGTTTCAGAA - - - - - TATGTAA	367
Db	133	GTCGTGCTATCTCGGACAGCACTGATGTTAAAGAGCTTCATAGTTTAAGAATTATCTAA	192
OY	368	AATATGTTTAAAAA CTGTAAAGCCGACGACGTGAATTTTACACCCAGTTACTAGAAAAAG	427
Db	193	AATATTTTAAAAA CTATAAAGCTGCACACATGATTTTACACCTAGTTACTAGAAAAGT	252
OY	428	AAGGGAAGCACTAGTCACTGAGTTAAAGGAAGCTGAAAACGAGAACGCACTTTACTATTC	487
Db	253	AAGGAAAAGCACTTATTAAGTCTGMAAA - - - - - AAGTAAACATAGGAAAGCACTTTACTAT	307
OY	488	TACCAAAAAAATCTCCGAATGCATTTTCAGAAAGATCTTATAGTACAGCTCAGACATATT	547
Db	308	CGACAAAAAAACCTTTCTAATGCATTTATCGAAAGATTTTATATATACAAAG - GAGGCATATT	366
OY	548	GCTCGTTTAAAGAGGGGGCTTAAAGAAAAAGCACTTGCTAAGTTAGACACTGTGAGATGG	607
Db	367	GCCTCAGTCAGAAAGGGGTCTATTAAGAAAAGCACTTACTAAGTTACGCACTTAACGAAACAA	428
OY	608	CCAGTTTAAATATAGCACTCAGAGCCCCCATCTGGGAGGAGACGACGAGGGGAAAGG	662
Db	427	CCAGTTTAAAGATTAATTAATGCCCAATTTTGGGAGAGCATGCGAGGTGTAAAG	481

```

RESULT 12
US-10-066-543-2964
; Sequence 2964, Application US/10066543
; Publication No. US20030087818A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Yugu
; APPLICANT: Pyle, Ruth A.
; APPLICANT: Xu, Jianshun
; APPLICANT: Indrias, Carol Joseph
; APPLICANT: Lodes, Michael J.
; APPLICANT: Secrist, Heather
; APPLICANT: Carter, Darick
; APPLICANT: Fanger, Gary R.
; APPLICANT: Smith, Carole L.
; APPLICANT: Durham, Margarita
; APPLICANT: Stolk, John A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.563
; CURRENT FILING DATE: 2002-01-31
; NUMBER OF SEQ ID NOS: 3417
; SOFTWARE: FASTSeq for Windows Version 4.0
; SEQ ID NO 2964
; LENGTH: 567
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 527
; OTHER INFORMATION: n = A,T,C or G
US-10-066-543-2964

Query Match      14.8%; Score 187.8; DB 14; Length 567;
Best Local Similarity 73.3%; Pred. No. 3.2e-41;
Matches 304; Conservative 0; Mismatches 87; Indels 24; Gaps 4;

OY 265 TACCGAAGCATGTGTCGCCGTGGTGAGTCTAGAGGGGGCGCATCATCATGAGA 324
DB 74 TACAGAGGCGTGTGTCCCGCGGCGAGTAAGA-GGGTCAGTATTGTGTCAGAGACA 132
OY 335 GTGTGTGTACTCTGCGAAGACAGTATGTTTCAGAA-----TATCTAA 367
DB 133 GTGTGTAATCTGTGCGAAGACAGTATGTTTAAAGTTTCAATGTTAATATCTAA 192
OY 368 AATAGTTTAAATACGTAAAGCCGACGCGATTTCTACACCCAGTTACTAGAAACG 427
DB 193 AATATTTTAAATACGTAAAGCCGACGCGATTTTACACCTAGTTACTAGAAACT 252
OY 428 AAGGAAGCAGTACGTAGTAAAGAGGTGAAACAGAACGCACTTCTACTATC 487
DB 253 AAGGAAGCAGTATTAAGTCTGTAATA-----AAGTAACATGAAAGCACTTTTACTAT 307
OY 488 TACCAAAAAAATCTCCGAATGATTATCAGAAAGATCTTATAGTACAGGTGACATATT 547
DB 308 CGACAAAAAACCCTTATATGATCATATCAGAAAGATTTTATATACAG-GAGGCATATT 366
OY 548 GTCCTTAAAGAGGGGGTCTTAAAGAAAGCACTTCTAGTTAGCACTGTGAGATGG 607
DB 367 GTCAGTACGAGAGGGGTCTTAAAGAAAGCACTTCTAGTTAGCACTGAGCAAGAA 426
OY 608 CCAGTTTAAATGAGCTCAAGCCCATCTGGGGAGGAGACAGAGGGGGAAGGG 662
DB 427 CCAGTTTAAAGATGATTAATGATCCCAATTTGGGAGGAGTGGCAGGTAAAGAG 481

```

```

; APPLICANT: ENDEGE, WILSON O., ET AL.
; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
; FILE REFERENCE: CCDA-260XX
; CURRENT FILING DATE: US/09/871,161
; PRIOR FILING DATE: 2001-05-31
; PRIOR APPLICATION NUMBER: 09/328,111
; PRIOR FILING DATE: 1999-06-08
; PRIOR APPLICATION NUMBER: 60/117,393
; PRIOR FILING DATE: 1999-01-27
; PRIOR APPLICATION NUMBER: 60/098,639
; NUMBER OF SEQ ID NOS: 544
; SOFTWARE: FASTSeq for Windows Version 3.0
; SEQ ID NO 483
; LENGTH: 605
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(605)
; OTHER INFORMATION: n = A,T,C or G
US-09-871-161-483

Query Match      14.8%; Score 187.2; DB 11; Length 605;
Best Local Similarity 65.5%; Pred. No. 4.8e-41;
Matches 348; Conservative 0; Mismatches 158; Indels 25; Gaps 5;

OY 265 TACCGAAGCATGTGTCGCCGTGGTGAGTCTAGAGGGGGCGCATCATCATGAGA 324
DB 74 TACAGAGGCGTGTGTCCCGCGGCGAGTAAGA-GGGTCAGTATTGTGTCAGAGACA 132
OY 325 GTGTGTGTACTCTGCGAAGACAGTATGTTTCAGAA-----TATCTAA 367
DB 133 GTGTGTAATCTGTGCGAAGACAGTATGTTTAAAGTTTCAATGTTAATATCTAA 192
OY 368 AATAGTTTAAATACGTAAAGCCGACGCGATTTCTACACCCAGTTACTAGAAACG 427
DB 193 AATATTTTAAATACGTAAAGCCGACGCGATTTTACACCTAGTTACTAGAAACT 252
OY 428 AAGGAAGCAGTACGTAGTAAAGAGGTGAAACAGAACGCACTTCTACTATC 487
DB 253 AAGGAAGCAGTATTAAGTCTGTAATA-----AAGTAACATGAAAGCACTTTTACTAT 307
OY 488 TACCAAAAAAATCTCCGAATGATTATCAGAAAGATCTTATAGTACAGGTGACATATT 547
DB 308 CGACAAAAAACCCTTATATGATCATITTCAGAAAGATTTTATATACAG-GAGGCATATT 366
OY 548 GTCCTTAAAGAGGGGGTCTTAAAGAAAGCACTTCTAGTTAGCACTGTGAGATGG 607
DB 367 GTCAGTACGAGAGGGGTCTTAAAGAAAGCACTTCTAGTTAGCACTGAGCAAGAA 426
OY 608 CCAGTTTAAATGAGCTCAAGCCCATCTGGGGAGGAGACAGAGGGGGAAGGG 667
DB 427 CCAGTTTAAAGATGATTAATGATCCCAATTTGGGAGGAGTGGTAAAGAAAN 486
OY 668 CAAGAGAGCAGTATGATGAGTCCCATTTGTC-ATCTAGTGTGACAGAAATTAACG 726
DB 487 GAAAGCAGTAAAGAAAGCACTTCTGTTATACAAACCTTTTATTCTACTAGTAT 546
OY 727 TTAAGAAAGCTTACCGGTGACACTTTTATTAAGTTAGTGAATTAACATGATAC 777
DB 547 TGACAAAAATTAACCTTTTAAAGTTTAAAGTTTAAAGTTTAAAGTTTAAAGTTT 597

```

```

RESULT 13
US-09-871-161-483
; Sequence 483, Application US/09871161
; Publication No. US20030097666A1
; GENERAL INFORMATION:

```

```

RESULT 14
US-10-066-543-3022
; Sequence 3022, Application US/10066543
; Publication No. US20030087818A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Yugu
; APPLICANT: Pyle, Ruth A.
; APPLICANT: Xu, Jianshun

```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 25, 2003, 21:56:10 : Search time 3752.19 Seconds

(without alignments)
11088.207 Million cell updates/sec

Title: US-09-717-321a-15

Perfect score: 1017

Sequence: 1 cccattctctgctcagatc.....ccttggtctgtaggttc 1017

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: gb_da:*
2: gb_hlg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_da:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vl:*
30: em_hlg_hum:*
31: em_hlg_inu:*
32: em_hlg_other:*
33: em_hlg_mus:*
34: em_hlg_pln:*
35: em_hlg_rtd:*
36: em_hlg_mam:*
37: em_hlg_vrt:*
38: em_sy:*
39: em_hlgo_hum:*
40: em_hlgo_mus:*
41: em_hlgo_other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1017	100.0	1017	6 AX163751	AX163751 Sequence
2	1017	100.0	1266	2 AX163738	AX163738 Sequence
3	1013.8	99.7	217700	2 AC106124	AC106124 Rattus no
4	679.2	66.8	2281	10 BC051053	BC051053 Mus muscu
5	677.6	66.6	2319	10 BC003828	BC003828 Mus muscu
6	677.6	66.6	269081	2 AC068493	AC068493 Mus muscu
7	403.2	39.6	185245	2 AC102775	AC102775 Mus muscu
8	403.2	39.6	202342	10 AL583884	AL583884 Mouse DNA
9	337.2	33.2	179685	10 AC126556	AC126556 Mus muscu
10	290.6	28.6	1022	6 AX163752	AX163752 Sequence
11	290.6	28.6	2302	9 BC050687	BC050687 Homo sapi
12	290.6	28.6	2315	9 AK054993	AK054993 Homo sapi
13	290.6	28.6	28567	9 HSA132695	AJ132695 Homo sapi
14	290.6	28.6	212827	9 AC009412	AC009412 Homo sapi
15	285.2	28.0	1232	6 AX408985	AX408985 Sequence
16	285.2	28.0	1232	9 HMP02579	D25274 Homo sapien
17	194.6	19.1	202565	9 AL354696	AL354696 Human DNA
18	193.2	19.0	605	6 AX341407	AX341407 Sequence
19	190.2	18.7	192498	2 AC105979	AC105979 Mus muscu
20	190.2	18.7	230015	2 AC132602	AC132602 Mus muscu
21	173	17.0	137625	9 AC104663	AC104663 Homo sapi
22	169.8	16.7	3544	9 AF542527	AF542527 Homo sapi
23	116.2	11.4	137625	9 AC104663	AC104663 Homo sapi
24	113.8	11.2	87616	2 AC139405	AC139405 Homo sapi
25	113	11.1	348	6 AX163753	AX163753 Sequence
26	112.4	11.1	64781	2 AC102025	AC102025 Mus muscu
27	105.8	10.4	101584	9 CNG01D85	AL121655 BAC sequ
28	105.8	10.4	155943	9 AC012364	AC012364 Homo sapi
29	103.4	10.2	5544	9 AF542527	AF542527 Homo sapi
30	103.2	10.1	110816	9 AC002404	AC002404 Human Chr
31	98.4	9.7	170839	2 AC133467	AC133467 Mus muscu
32	85.6	8.4	495	6 AX341256	AX341256 Sequence
33	84.2	8.3	180303	9 AL672045	AL672045 Human DNA
34	84.2	8.3	201012	2 AC021189	AC021189 Homo sapi
35	76.2	7.5	228121	2 AC133022	AC133022 Rattus no
36	76.2	7.5	239768	2 AC112582	AC112582 Rattus no
37	76.2	7.5	245468	2 AC130746	AC130746 Rattus no
38	67	6.6	594	6 AX385363	AX385363 Sequence
39	62.8	6.2	194371	2 AC135453	AC135453 Rattus no
40	61.4	6.0	378	6 BD030694	BD030694 Sequence
41	56.8	5.6	198	11 G31747	G31747 sWSS2749 Er
42	54.2	5.3	7218	6 I66494	I66494 Sequence 14
43	52.4	5.2	162	6 AX397334	AX397334 Sequence
44	51.4	5.1	90354	9 AP001124	AP001124 Homo sapi
45	51.4	5.1	178089	2 AC010929	AC010929 Homo sapi

ALIGNMENTS

RESULT 1
AX163751
LOCUS AX163751 1017 bp DNA linear PAT 22-JUN-2001
DEFINITION Sequence 15 from Patent WO0138579.
ACCESSION AX163751
VERSION AX163751.1 GI:14544857
KEYWORDS
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1
Gould-Rothberg,B.E., Dippio,V.A., Ramsehl,T.M. and Gerweil,R.W.
Method of identifying toxic agents using nsaid-induced differential

gene expression in liver
Patent: WO 0138579-A 15 31-MAY-2001;
Curegen Corporation (US)
Location/Qualifiers
1. .1017
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"

BASE COUNT 245 a 245 c 216 g 311 t
ORIGIN

Query Match 100.0%; Score 1017; DB 6; Length 1017;
Best Local Similarity 100.0%; Pred. No.1.2e-291;
Matches 1017; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCCCTATTCTTGGCTCAGATTAAAGATTGCCAAAATACCTGTGTGAAGTAAAGTTGGCTGTG 60
Db 1 CCCCTATTCTTGGCTCAGATTAAAGATTGCCAAAATACCTGTGTGAAGTAAAGTTGGCTGTG 60

QY 61 CTGGAACAGCTAGACACTAAGCTGTGAGAGACTTTGTCTTAAAGAGACTGCAGCTTC 120
Db 61 CTGGAACAGCTAGACACTAAGCTGTGAGAGACTTTGTCTTAAAGAGACTGCAGCTTC 120

QY 121 TGGGCTCAGGGGTGCAAGCCCTCCCGTAGCTCCAGACCGGTGACACAGACAGCCCTCC 180
Db 121 TGGGCTCAGGGGTGCAAGCCCTCCCGTAGCTCCAGACCGGTGACACAGACAGCCCTCC 180

QY 181 TTAATGACAGCTGCCATGTAAAGCCACCTGTAATCACTATACAGCCGATCATACGTAAAC 240
Db 181 TTAATGACAGCTGCCATGTAAAGCCACCTGTAATCACTATACAGCCGATCATACGTAAAC 240

QY 241 TTTGTACTGTACGTCAAGATGGGTGTAAAGAGCTGCTGCTTGTGATTTCAATAGTAGTTCT 300
Db 241 TTTGTACTGTACGTCAAGATGGGTGTAAAGAGCTGCTGCTTGTGATTTCAATAGTAGTTCT 300

QY 301 CTAAATATCAGACGTGACCGGGCTTCTGCAGGCTTTGAACAGAACTCTGGCTCTGTGTGC 360
Db 301 CTAAATATCAGACGTGACCGGGCTTCTGCAGGCTTTGAACAGAACTCTGGCTCTGTGTGC 360

QY 361 CTCTAACGAAGTATTCGTTCTTAAGTGTGGGTGTGCTGGGTGGAAGTGTGAAACACGA 420
Db 361 CTCTAACGAAGTATTCGTTCTTAAGTGTGGGTGTGCTGGGTGGAAGTGTGAAACACGA 420

QY 421 CGTCATCAAGGAGACAGACGATTTTGAATATGAAGTAAAGTAACTTAATTACACTA 480
Db 421 CGTCATCAAGGAGACAGACGATTTTGAATATGAAGTAAAGTAACTTAATTACACTA 480

QY 481 CATTTGACATGAGATAATTCMACTGAATAAAGTGTACAGGGTAAAGCTTTTAAAGGTT 540
Db 481 CATTTGACATGAGATAATTCMACTGAATAAAGTGTACAGGGTAAAGCTTTTAAAGGTT 540

QY 541 AATTTCTGTCAACAGATAGATGACAAATGGCCGATCTTAATCAGTGTCTCTTGAGCCCC 600
Db 541 AATTTCTGTCAACAGATAGATGACAAATGGCCGATCTTAATCAGTGTCTCTTGAGCCCC 600

QY 601 CCTTCCCCCTGCTGCTCCCTCCCGATGGGGCTGTGATCATTTTAACTGGCCATCC 660
Db 601 CCTTCCCCCTGCTGCTCCCTCCCGATGGGGCTGTGATCATTTTAACTGGCCATCC 660

QY 661 TCACAGTTGCTAACTTGAAGTGTCTTTTGAAGACCCCTCTTAAGAGCAATATG 720
Db 661 TCACAGTTGCTAACTTGAAGTGTCTTTTGAAGACCCCTCTTAAGAGCAATATG 720

QY 721 TCTGACCTGTACTATAAGATCTTTCTATAATGCAATCGAGATTTTGTGGTAGATAGT 780
Db 721 TCTGACCTGTACTATAAGATCTTTCTATAATGCAATCGAGATTTTGTGGTAGATAGT 780

QY 781 AGAAGTGTCTTCCGTTTACCTTCTTACTAGTGAAGTACTAGTCTCCCTCGTTT 840
Db 781 AGAAGTGTCTTCCGTTTACCTTCTTACTAGTGAAGTACTAGTCTCCCTCGTTT 840

QY 841 CTAGTAAGTGGGTAGAAATCAAGTGTGCGGCTTACAGTTTAACTAATTTAGAT 900
Db 841 CTAGTAAGTGGGTAGAAATCAAGTGTGCGGCTTACAGTTTAACTAATTTAGAT 900

Db 841 CTAGTAAGTGGGTAGAAATCAAGTGTGCGGCTTACAGTTTAACTAATTTAGAT 900

QY 901 ATTCTGAACATCAGCTGTCTTGGCCAGAGTACCAACACTGTATGATGATGCGGCC 960
Db 901 ATTCTGAACATCAGCTGTCTTGGCCAGAGTACCAACACTGTATGATGATGCGGCC 960

QY 961 CTCTAGACCTCACCACCGGAGACATAGCTTCCGGTACCTTTGGGTGTGAGGTTG 1017
Db 961 CTCTAGACCTCACCACCGGAGACATAGCTTCCGGTACCTTTGGGTGTGAGGTTG 1017

RESULT 2
AXI63738/c
LOCUS AXI63738 1266 bp DNA Linear PAT 22-JUN-2001
DEFINITION Sequence 2 from Patent WO0138579.
VERSION AXI63738.1 GI:14544844
KEYWORDS
SOURCE
ORGANISM
Rattus norvegicus (Norway rat)
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE
1
AUTHORS Gould-Rothberg B.E., DiIppio V.A., Ramseh, T.M. and Gerwein, R.W.
TITLE Method of Identifying Toxic Agents using nsald-Induced differential
gene expression in liver
Patent: WO 0138579-A 2 31-MAY-2001;
Curegen Corporation (US)
LOCATION/Qualifiers
1. .1266
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"

BASE COUNT 385 a 258 c 285 g 338 t
ORIGIN

Query Match 100.0%; Score 1017; DB 6; Length 1266;
Best Local Similarity 100.0%; Pred. No.1.3e-291;
Matches 1017; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCCCTATTCTTGGCTCAGATTAAAGATTGCCAAAATACCTGTGTGAAGTAAAGTTGGCTGTG 60
Db 1261 CCCCTATTCTTGGCTCAGATTAAAGATTGCCAAAATACCTGTGTGAAGTAAAGTTGGCTGTG 1202

QY 61 CTGGAACAGCTAGACACTAAGCTGTGAGAGACTTTGTCTTAAAGAGACTGCAGCTTC 120
Db 1201 CTGGAACAGCTAGACACTAAGCTGTGAGAGACTTTGTCTTAAAGAGACTGCAGCTTC 1142

QY 121 TGGGCTCAGGGGTGCAAGCCCTCCCGTAGCTCCAGACCGGTGACACAGACAGCCCTCC 180
Db 1141 TGGGCTCAGGGGTGCAAGCCCTCCCGTAGCTCCAGACCGGTGACACAGACAGCCCTCC 1082

QY 181 TTAATGACAGCTGCCATGTAAAGCCACCTGTAATCACTATACAGCCGATCATACGTAAAC 240
Db 1081 TTAATGACAGCTGCCATGTAAAGCCACCTGTAATCACTATACAGCCGATCATACGTAAAC 1022

QY 241 TTTGTACTGTACGTCAAGATGGGTGTAAAGAGCTGCTCTTGTGATTTCAATAGTAGTTCT 300
Db 1021 TTTGTACTGTACGTCAAGATGGGTGTAAAGAGCTGCTCTTGTGATTTCAATAGTAGTTCT 962

QY 301 CTAAATATCAGACGTGACCGGGCTTCTGCAGGCTTTGAACAAACTCTGGCTCTGTGTGC 360
Db 961 CTAAATATCAGACGTGACCGGGCTTCTGCAGGCTTTGAACAAACTCTGGCTCTGTGTGC 902

QY 361 CTCTAACGAAGTATTCGTTCTTAAGTGTGGGTGTGCTGGGTGGAAGTGTGAAACACGA 420
Db 901 CTCTAACGAAGTATTCGTTCTTAAGTGTGGGTGTGCTGGGTGGAAGTGTGAAACACGA 842

QY 421 CGTCATCAAGGAGACAGACGATTTTGAATATGAAGTAAAGTAACTAATTTAGACTA 480
Db 841 CGTCATCAAGGAGACAGACGATTTTGAATATGAAGTAAAGTAACTAATTTAGACTA 782

* consists of 11 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 109797: contig of 109797 bp in length
* 109798 109897: gap of unknown length
* 109898 166934: contig of 57037 bp in length
* 166935 167034: gap of unknown length
* 167035 176384: contig of 9350 bp in length
* 176385 176484: gap of unknown length
* 176485 206327: contig of 29843 bp in length
* 206328 206427: gap of unknown length
* 206428 207461: contig of 1034 bp in length
* 207462 208607: gap of unknown length
* 208608 208707: contig of 1046 bp in length
* 208708 209908: gap of unknown length
* 209909 210008: gap of unknown length
* 210009 211583: contig of 1575 bp in length
* 211584 211683: gap of unknown length
* 211684 213649: contig of 1966 bp in length
* 213650 213748: gap of unknown length
* 213750 215115: contig of 1366 bp in length
* 215116 215215: gap of unknown length
* 215216 217700: contig of 2485 bp in length.

FEATURES

SOURCE

1. 217700
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-119E11"

BASE COUNT 51384 a 45640 c 45954 g 51476 t 23246 others
ORIGIN

Query Match 99.7%: Score 1013.8; DB 2; Length 217700;
Best Local Similarity 99.88; Pred. No. 1.7e-290;
Matches 1015; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CCCCTATTCTTGTCTGACATTAAGAAATGGCCAAATACCTTGTGAAGTGGCTGTG 60
Db 63447 CCCCTATTCTTGTCTGACATTAAGAAATGGCCAAATACCTTGTGAAGTGGCTGTG 63506
QY 61 CTGGAACACGTAAAGCACTAAGCTGTGAGAGACTTGTCTTAAAGAACTGCAGCTTC 120
Db 63507 CTGGAACACGTAAAGCACTAAGCTGTGAGAGACTTGTCTTAAAGAACTGCAGCTTC 63566
QY 121 TGGGCTGAGGGGTGACAGCCCTCCGTAAGTCCAGACCCGTTGACACAGCAGCCCTCC 180
Db 63567 TGGGCTGAGGGGTGACAGCCCTCCGTAAGTCCAGACCCGTTGACACAGCAGCCCTCC 63626
QY 181 TTAATGACAGCTGCCATGTAAACGACCTGTAACTTATACAGCCCATCTCTATAGTAAC 240
Db 63627 TTAATGACAGCTGCCATGTAAACGACCTGTAACTTATACAGCCCATCTCTATAGTAAC 63686
QY 241 TTGTCTACTGTACGTACAGATGGGTGTAACAGCTCTGCTTTTGAATTCATAGTAGTTC 300
Db 63687 TTGTCTACTGTACGTACAGATGGGTGTAACAGCTCTGCTTTTGAATTCATAGTAGTTC 63746
QY 301 CTAAATATACAGCTGACCGGCTTCTGACAGCTTTGAACAGACCTGCTCTCTGTGTC 360
Db 63747 CTAAATATACAGCTGACCGGCTTCTGACAGCTTTGAACAGACCTGCTCTCTGTGTC 63806
QY 361 CTCTAAGCAAGTATTTCTTCTTCTAGTGTGGGTGTGGGTGGGTGGGTGGTGAACACGA 420
Db 63807 CTCTAAGCAAGTATTTCTTCTTCTAGTGTGGGTGTGGGTGGGTGGGTGGTGAACACGA 63866
QY 421 CGTCATCAAAAGAGACAGACAGATATTTTGAATAATAGAACTAGAGATTAATTACACTA 480
Db 63867 CGTCATCAAAAGAGAGACAGACAGATATTTTGAATAATAGAACTAGAGATTAATTACACTA 63926

QY 481 CATTGTACATGAGACTAATTAACCTGAATTAAGTGCACGGGTAAAGCTTTTAAACGGTT 540
Db 63927 CATTGTACATGAGACTAATTAACCTGAATTAAGTGCACGGGTAAAGCTTTTAAACGGTT 63986
QY 541 AATTCTGTCAACAGATAGATGACAAATGGCCGATCTTATACAGTCTCTCTGAGGCC 600
Db 63987 AATTCTGTCAACAGATAGATGACAAATGGCCGATCTTATACAGTCTCTCTGAGGCC 64046
QY 601 CCTTCCCTGCTGCTGCTCCCTCCAGATGGGCGGTGAGTCCATATTTAACTGGCCATCC 660
Db 64047 CCTTCCCTGCTGCTGCTCCCTCCAGATGGGCGGTGAGTCCATATTTAACTGGCCATCC 64106
QY 661 TCACAGTTGCTACTTACGAAATGCTCTTTTGAAGACCCCTTTTAAAGCAGCAATATG 720
Db 64107 TCACAGTTGCTACTTACGAAATGCTCTTTTGAAGACCCCTTTTAAAGCAGCAATATG 64166
QY 721 TCTGACGTGATTAAGATCTTCTGATATATGATCGGAGATTTTGGTGGATAGT 780
Db 64167 TCTGACGTGATTAAGATCTTCTGATATATGATCGGAGATTTTGGTGGATAGT 64226
QY 781 AGAAGTCCGTCTGCTTTCACCTTCTTACTGCTGACTAGTGTCCCTTCTGCTTT 840
Db 64227 AGAAGTCCGTCTGCTTTCACCTTCTTACTGCTGACTAGTGTCCCTTCTGCTTT 64286
QY 841 CTACTACTGGGTGTAGAAATACAGTCTGCTGCGCTTTTAACTTTTAACTATTTAGAT 900
Db 64287 CTACTACTGGGTGTAGAAATACAGTCTGCTGCGCTTTTAACTTTTAACTATTTAGAT 64346
QY 901 ATTCTGAACATCACTGCTCTTCCAGATGACCACTGATCATGTGATGATGCGGCC 960
Db 64347 ATTCTGAACATCACTGCTCTTCCAGATGACCACTGATCATGTGATGATGCGGCC 64406
QY 961 CTCTAGACCTTACCCACGCGGACACATGCTTCCGTAACCTTTGGGTGTGAGGTTTC 1017
Db 64407 CTCTAGACCTTACCCACGCGGACACATGCTTCCGTAACCTTTGGGTGTGAGGTTTC 64463

RESULT 4

BC051053

DEFINITION Mus musculus, clone MGC:58966 IMAGE:508182, mRNA, complete cds.
ACCESSION BC051053
VERSION BC051053.1 GI:29835221
KEYWORDS MGC.

SOURCE

ORGANISM

MUS MUSCULUS

EUKARYOTA; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE

1 (bases 1 to 2281)

AUTHORS

Strausberg, R.

TITLE

Direct Submission

JOURNAL

COMMENT

REMARK

NIH-MGC Project URL: <http://mgc.ncl.nih.gov>
Contact: MGC help desk
Email: cgabs-remail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLN)
DNA Sequencing by: National Institutes of Health Intramural Sequencing Center (NISC),
Gaithersburg, Maryland;
Web site: <http://www.nisc.nih.gov/>
Contact: nisc.mgc@nih.gov
Ahter, N., Ayle, K., Beckstrom-Sternberg, S.M., Benjamin, B.,
Blakesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S.,
Diehl, N.L., Grantham, S., Guan, X., Gupta, J., Haghighi, P.,
Hansen, N., Ho, S.-L., Karins, E., Kwong, P., Lalic, P., Legaspi, R.,
Maduro, O.L., Masello, C., Maskell, B., Mastrian, S.D., McCloskey, J.C.,
McGowan, J., Pearson, R., Stantrop, S., Thomas, P.J., Touchman, J.W.,
Tsurgoun, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L.,

Young, A., Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/BLIN at: <http://image.llnl.gov>
Series: IRAX Plate: 108 Row: f Column: 9
This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis, Genomescan gene prediction, Similarity but not identity to protein.

FEATURES
source Location/Qualifiers

1..2281
/organism="Mus musculus"
/mol_type="mRNA"
/strain="Czech 11"
/db_xref="taxon:10090"
/clone="MGC:58966 IMAGE:5038182"
/issue_type="Mammary tumor metastasized to lung. Tumor arose spontaneously from a senescent normal mammary (clonal) outgrowth infected with the virus MMTV."
/clone_1lb="NCI_CGAP_1429"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6"
197..775
/codon_start=1
/product="Unknown (protein for MGC:58966)"
/protein_id="AAH51053.1"
/db_xref="GI:29835222"
/translation="MOAICVYVGDGAVGKTCGLISYTNAPPEXIPVDFNYSANV
MYDKRVNLGLMDTQEDYDLRLPISOTYVELICFSLSPASSENRAKPIYPRV
HHCNPPIIVSTKIDLRDQKTIKAKKILPTITPPOGLAMAKETGAVKILCSAL
TORGKTVDEIRAVLCPPIKRRKKRLLL"
BASE COUNT 568 a 569 c 504 g 640 t
ORIGIN

Query Match 66.8%; Score 679.2; DB 10; Length 2281;
Best Local Similarity 86.1%; Pred. No. 6,7e-191;

Matches 876; Conservative 0; Mismatches 113; Indels 29; Gaps 10;

1 CCCCTATTCCTGCTCAGATTAGAAATGCGAAATACCTGTGTAAGTTCGTTG 60
1019 CCCCATCTCTGCTGATTAATAAATGCGAAATACCTGTGTAAGTTCGTTG 1078
61 CTGAGAACAGCTAAGCAGTGTGAGAGA--CTTGTCTCTTAAGACAGCTGCAC 117
1079 CTGAGAACAGCTAAGCAGTGTGAGAGA--CTTGTCTCTTAAGACAGCTGCAC 1138
118 TTCTGGGCTCAGGGGTCAGACCCCTCCAGTCTCCAGACCGGTGACACAGCAGCC 177
1139 TTCTGGAGTCAGGGGTCAGACCCCTCCAGTCTCCAGACCGGTGACACAGCAGCC 1198
178 TCCTTAATGACAGCGCTGCATGTACGACAGCTGTATATCAGCCATGCTCATTCG 237
1199 TCCTTAATGACAGCGTGCATGTACGACAGCTGTATATCAGCCATGCTCATTCG 1258
238 AACTTTGTACTGACTGCAC--GATGGGTGTACAGCTCTGCTCTTTGATTTCAATG 294
1259 AACTTTGTACTGACTGCACAGTGTGACAGCTGCTCTTTGATTTCAATG 1318
295 AACTTCTTAATTAATACAGCTGCACCGCTCTGAGGCTTGAACAGACAGCTGCCTG 354
1319 AACTTTTCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1378
355 TGTGCTCTTAAGAGATTTCTGTTCTAGTCTGAGGTGTGCTGAGGTGAGTGTGA 414
1379 TGTGCTCTTAAGAGATTTCTGTTCTAGTCTGAGGTGTGCTGAGGTGAGTGTGA 1436
415 AACAGCAGTCATCAAAAGAGACAGACAGATTTTGAATATGAATGAATGAATTAAT 474
1437 AACAGCAGTCATCAAAAGAGACAGACAGATTTTGAATATGAATGAATGAATTAAT 1496
475 AACAGCAGTCATCAAAAGAGACAGACAGATTTTGAATATGAATGAATGAATTAAT 534
1497 AACAGCAGTCATCAAAAGAGACAGACAGATTTTGAATATGAATGAATGAATTAAT 1556

QY 535 ACGTTAATTTCTGTCAAACAGTAGATGACAATAGGCGCATTTATCAGTGTCTC----- 589
DB 1557 ACGTTAATTTCTGTCAAACAGTAGATGACAATAGGCGCATTTATCAGTGTCTC 1616
QY 590 -----TCTGAGACCCCCCTCCCTCCCTGCTGTCCTCCAGATGGGCGTTAGTCA 642
DB 1617 CATCCCTCGCGCCGCCACCCCTCCAGTCCGCTCCCTCCAGATGGGCGTTAGTCA 1676
QY 643 TATTTAAGTCGCGCCGCCACCCCTCCAGTCCGCTCCCTCCAGATGGGCGTTAGTCA 702
DB 1677 TATTTAAGTCGCGCCGCCACCCCTCCAGTCCGCTCCCTCCAGATGGGCGTTAGTCA 1734
QY 703 TTCTTAACAGCAGATATGTCGACCTGTACTATAGATCTTCTGATTAATGATTGCG- 761
DB 1735 TTCTTAACAGCAGATATGTCGACCTGTACTATAGATCTTCTGATTAATGATTGCG 1794
QY 762 GATTTTGTGTAAGATGTAAGAGTGGCTTCTGTTTACCTTCTTACTCAG-CTGA 820
DB 1795 TTTTGTGTAAGATGTAAGAGTGGCTTCTGTTTACCTTCTTACTCAGCTGA 1854
QY 821 CTAGTGTCTCTGTTCTGTAAGTGGGTGTAAGATGAGTGGCGCTTACTA 880
DB 1855 GTTAGTGTCTCTGTTCTGTAAGTGGGTGTAAGATGAGTGGCGCTTACTA 1914
QY 881 GTTTTAACTATTTAGATATTCGTAAC--ATCAGTGTCTTCCAGAGTACCAACT 938
DB 1915 GTTTTAACTATTTAGATATTCGTAAC--ATCAGTGTCTTCCAGAGTACCAACT 1974
QY 939 GTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 996
DB 1975 GTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2029

RESULT 5
BC003828

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

1 (bases 1 to 2319)

Strausberg, R.L., Feilung, E.A., Grouse, L.H., Derge, J.G.,

Klausner, R.D., Collins, F.S., Wagner, L., Shennan, C.M., Schuler, G.D.,

Altschul, S.F., Zeeberg, B., Buettow, K.H., Schaefer, C.F., Bhat, N.K.,

Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F.,

Datchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,

Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,

Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S.,

Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J.,

Abrahamson, R.D., Mullany, S.J., Bosak, S.A., McEwan, P.J.,

McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,

Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Huliy, S.W.,

Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,
Sanchez, J., Helton, E., Kettelman, M., Madan, A.C., Rodrigues, S.,
Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y.,
Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D.,
Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmitz, J., Myers, R.M.,
Butterfield, Y.S., Kravitski, M.I., Skalski, U., Small, D.E.,
Schnerch, A., Schein, J.E., Jones, S.J. and Marra, M.A.
Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

JOURNAL

MEDLINE

PUBMED

2 (bases 1 to 2319)

Strausberg, R.

REFERENCE

AUTHORS

TITLE

Direct Submission

JOURNAL Submitted (28-FEB-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

REMARK NIH-MGC Project URL: <http://mgc.ncl.nih.gov>

COMMENT Contact: MGC help desk
Email: egapds-remail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
Center, Stanford University School of Medicine, Stanford, CA 94305
Web site: <http://www-shgc.stanford.edu>
Contact: (Dickson, Mark) mcdepaxil.stanford.edu
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.

FEATURES
source Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAC Plate: 11 Row: e Column: 22.
Location/Qualifiers
1. .2319
/organism="Mus musculus"
/mol_type="mRNA"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="MGC:6235 IMAGE:3593957"
/tissue_type="Mammary tumor. Metastatic adenoma-tgf alpha model. 10 month old virgin mouse. Taken by biopsy."
/clone_11b="NCI CGAP_Mam1"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6"
1. .2319
/gene="Rac1"
/note="synonym: D5Erlc559e"
/db_xref="locusid:19353"
/db_xref="MGI:97845"
198..776
/codon_start=1
/product="Rac1 protein"
/protein_id="AAH03828.1"
/db_xref="GI:13277918"
/db_xref="locusid:19353"
/translation="MOAIKCYVVDGAVGKCLISTTTNAPFEYIPYVDNSANV
HVDKPYNLGMDTAGEBDYDLRLPLSTPQDVLFCISPSFSEVVRKWPPEVR
HHCNPTILLVGTGLDLRDDDTLEKLEKLTPTITYPQGLAMKEIGAVYLECSAL
TORGLKTFEFAIRAVLCPPVKKRKRCLL."

CDS
BASE COUNT 603 a 573 c 503 g 640 t
ORIGIN

Query Match 66.6%; Score 677.6; DB 10; Length 2319;
Best Local Similarity 85.3%; Pred. No. 2e-190;
Matches 870; Conservative 0; Mismatches 119; Indels 31; Gaps 9;

OY 1 CCCCTATCTCTGCACAGATTAAAGATTGCCAAATACCTTGTAAGTAAGTGGCTTG 60
DB 1020 CCCCATTCTGTCACAGTATTAATAATGCAAAATACCTTGTAAGTAAGTGGCTTG 1079
OY 61 CTGGAACACGTAAAGCTAAGCTGTGAGAGA---CTTGTCTTAAGAAAGACTGCAGC 117
DB 1080 CTGGAACACCTTAAGCTAAGCTGTGAGAGACTCTGTGCTTAAGAAAGACTGCAGC 1139
OY 118 TTTCGGGCTCAGGGGCTCAGACCTCCCGTAGCTCCGAGCCGTGTGACACAGCAGCC 177
DB 1140 TTTCGGAGCCAGGGGCTCAGACCTCTTCAGAGTCCGAGCCGTGTGACACAGCAGCC 1199
OY 178 TCCTTAATGACACCTGCTGATGACGACCTGTAACTTATCAGCCCATGCTCAATAGCT 237
DB 1200 TCCTTAATGACACCTGCTGATGACGACCTGTAACTTATCAGCCCATGCTCAATAGCT 1239
OY 238 AACCTTGTACTGATGATCAGC---GATGGGTGTAACAGCTGTGCTTTGATTTAGTATG 294
DB 1260 AACCTTGTACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1319

OY 295 AGTTCTCTAAATACACAGCTGACCGGCTTCGACAGCTTTCAGACAACTCTGCTCTG 354
DB 1320 AGTTCTCTAAATACACAGCTGACCGGCTTCGACAGCTTTCAGACAACTCTGCTCTG 1379
OY 355 TGTTCCTCTAACAGATATTTCTTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTG 414
DB 1380 TGTTCCTCTAACAGATATTTCTTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTG 1437
OY 415 ACAGAGCTCATCAAGAGAGACAGAGATATTTGATTAATGATAGATATATTT 474
DB 1438 ACAGAGATGATCAAGAGAGACAGAGATATTTGATTAATGATAGATATATTT 1497
OY 475 ACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 534
DB 1498 ACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1557
OY 535 ACGGTTATATTTCTGCAAAAGTATGATGACAAATGCGCGATCTTATGAGTCTC----- 589
DB 1558 ACGGTTATATTTCTGCAAAAGTATGATGACAAATGCGCGATCTTATGAGTCTGAG 1617
OY 590 -----TCTTGAGCCGCCCTTCCCGCTGCTGCTGCTGCTGCTGCTGCTGCTG 642
DB 1618 CATGCCCTTCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCC 1677
OY 643 TATTTAACTGCGCATCCTCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 702
DB 1678 TATTTAACTGCGCATCCTCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1735
OY 703 TTTCTTAACGACATATGCTGACCTGACCTGACCTGACCTGACCTGACCTGACCTG 757
DB 1736 TTTCTTAACGACATATGCTGACCTGACCTGACCTGACCTGACCTGACCTGACCTG 1795
OY 758 CGGAGATTTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 816
DB 1796 TTTGTTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1855
OY 817 CTGACATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 876
DB 1856 GCTGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1915
OY 877 TACGATTTTAACTATTTTAACTATTTTAACTATTTTAACTATTTTAACTATTTTAA 936
DB 1916 TACGATTTTAACTATTTTAACTATTTTAACTATTTTAACTATTTTAACTATTTTAA 1975
OY 937 CTGTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 996
DB 1976 CTGTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2032

RESULT 6
AC068493 269081 bp DNA linear HTG 15-MAY-2002
LOCUS Mus musculus clone RP23-76k1 strain C57BL/6J, WORKING DRAFT
DEFINITION AC068493
AC068493
AC068493.10 GI:15148081
VERSION HTG_PHASE1; HTGS_DRAFT.
KEYWORDS Mus musculus (house mouse)
SOURCE Mus musculus
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 269081)
Montgomery,K.T., Grills,G., Han,J., Lee,E., Long,J., Pomerantz,R.,
Ioshikhes,I.P., Shlm,C., Decker,J., Thomas,E., Perera,A.,
Gordon,M., Goltz,J.S. and Kucherlapati,R.
High throughput Mouse Sequencing
TITLE
JOURNAL
REFERENCE
AUTHORS
MONTGOMERY,K.T., GRILLS,G., HAN,J., LEE,E., LONG,J., POMERANTZ,R.,
IOSHIKHES,I.P., SHLM,C., DECKER,J., THOMAS,E., PERERA,A.,
GORDON,M., GOLTZ,J.S. and KUCHERLAPATI,R.
Direct Submission

JOURNAL

COMMENT

Submitted (03-MAY-2000) Department of Molecular Genetics, Albert Einstein College of Medicine Genome Center, 1300 Morris Park Ave., Bronx, NY 10461, USA
On Aug 11, 2001 this sequence version replaced gl:14993654.

Center: Harvard Partners Genome Center

Center Code: HPGC

Web site: <http://www.hpcgg.org/Sequence/mouse.html>
Contact: hpcgcmendel.mgh.harvard.edu

-----Summary Statistics

Center project name: ABN

Sequencing vector: pUC18, L08752

Chemistry: Dye-terminator Big Dye; 100%

*Consensus quality: 256302 at least Q20

*Consensus quality: 254124 at least Q30

*Consensus quality: 250276 at least Q40

Estimated insert size: agarose-FP - N/A

*Estimated insert size: 268581 - sum-of-contigs

Quality coverage: agarose-FP - N/A

Quality coverage: 6.2 x in Q20 bases; sum-of-contigs estimation

* NOTE: This is a 'working draft' sequence. It currently

* consists of 26 contigs. The true order of the pieces

* is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence

* as soon as it is available and the accession number will

* be preserved

* 1 51730: contig of 51730 bp in length

* 51731 51750: gap of unknown length

* 51751 82423: contig of 30673 bp in length

* 82424 82444: gap of unknown length

* 118159 118158: contig of 35715 bp in length

* 118179 139441: gap of unknown length

* 139442 139461: contig of 21263 bp in length

* 139462 160638: contig of 21177 bp in length

* 160639 160658: gap of unknown length

* 160659 177611: contig of 16953 bp in length

* 177612 177631: gap of unknown length

* 177632 180991: contig of 13360 bp in length

* 180992 191011: gap of unknown length

* 191012 200128: contig of 9117 bp in length

* 200129 200148: gap of unknown length

* 200149 211882: contig of 11734 bp in length

* 211883 211902: gap of unknown length

* 211903 220773: contig of 8871 bp in length

* 220774 220793: gap of unknown length

* 220794 229643: contig of 8850 bp in length

* 229644 229663: gap of unknown length

* 229664 238339: contig of 8676 bp in length

* 238340 238360: gap of unknown length

* 238361 244848: contig of 6489 bp in length

* 244849 244868: gap of unknown length

* 244869 249725: contig of 4857 bp in length

* 249726 249745: gap of unknown length

* 249746 252523: contig of 2778 bp in length

* 252524 252543: gap of unknown length

* 252544 257520: contig of 4977 bp in length

* 257521 257541: gap of unknown length

* 257542 260396: contig of 2856 bp in length

* 260397 260416: gap of unknown length

* 260417 262709: contig of 2293 bp in length

* 262710 262729: gap of unknown length

* 262730 263934: contig of 1205 bp in length

* 263935 263954: gap of unknown length

* 263955 264196: contig of 242 bp in length

* 264197 264216: gap of unknown length

* 264217 264629: contig of 413 bp in length

* 264630 265608: gap of unknown length

* 265609 265628: gap of unknown length

* 265629 266935: contig of 1307 bp in length

* 266936 266955: gap of unknown length

* 266956 268335: contig of 1380 bp in length

* 268336 268355: gap of unknown length

* 268356 268549: contig of 194 bp in length

* 268550 268569: gap of unknown length

* 268570 269081: contig of 512 bp in length.

location/Qualifiers

1. 269081

FEATURES

source

/organism="Mus musculus"

/mol_type="genomic DNA"

/strain="C57BL/6/J"

/db_xref="taxon:10090"

/clone="RP23-76K1"

/sex="male"

1. 51730

/note="assembly_name:Contig138"

51751. 82423

/note="assembly_name:Contig137"

82444. 118158

/note="assembly_name:Contig136"

118179. 139441

/note="assembly_name:Contig135"

139462. 160638

/note="assembly_name:Contig134"

160659. 177611

/note="assembly_name:Contig133"

177632. 180991

/note="assembly_name:Contig132"

191012. 200128

/note="assembly_name:Contig131"

200149. 211882

/note="assembly_name:Contig130"

211903. 220773

/note="assembly_name:Contig129"

220794. 229643

/note="assembly_name:Contig128"

229664. 238339

/note="assembly_name:Contig127"

238360. 244848

/note="assembly_name:Contig126"

244869. 249725

/note="assembly_name:Contig125"

249746. 252523

/note="assembly_name:Contig124"

clone_end:SP6

vector_side:left"

252544. 257520

/note="assembly_name:Contig123"

257541. 260396

/note="assembly_name:Contig122"

260417. 262709

/note="assembly_name:Contig121"

262730. 263934

/note="assembly_name:Contig120"

263955. 264196

/note="assembly_name:Contig119"

264217. 264629

/note="assembly_name:Contig118"

264650. 265608

/note="assembly_name:Contig117"

265629. 266935

/note="assembly_name:Contig116"

266956. 268335

/note="assembly_name:Contig115"

268356. 268549

/note="assembly_name:Contig114"

268570. 269081

/note="assembly_name:Contig113"

BASE COUNT 69726 a 66027 c 64672 g 67724 t 932 others

ORIGIN

Query Match 66.6%; Score 677.6; DB 2; length 269081;

Best Local Similarity 85.38; Pred. No. 3e-190;
Matches 870; Conservative 0; Mismatches 119; Indels 31; Gaps 9;

```
QY 1 CCCCTATTCCTGTCAGATTAAAGATTGCCAAATACCTTGTGAAGTATGCGGTGTG 60
Db 62207 CCCCCTATTCCTGTCAGATTAAATATGTCACAAATACCTTGTGAAGTATGCGGTGTG 62206
QY 61 CTGGAACACGTAAGCACTAAGCTGTGAGAGA---CTTGTCTCTTAAGAAGCTGCAC 117
Db 62267 CTGGAACACGTAAGCACTAAGCTGTGAGAGACTTCTTGTCTTAAGAAGACCGGACC 62326
QY 118 TTTGGGGCTCAGGGGTGACGCTCCCTGATGCTCCAGACCGGTGACAGACAGCAGCC 177
Db 62327 TTTGGAGCCAGGGGTGACGCTCCCTGATGCTCCAGAGTCCAGACCGGTGACAGACAGCC 62386
QY 178 TCCCTAATGACACGCTGCATGTAAGCAGCCTGTACTATATCAGCCCATGCTCATTA 237
Db 62387 TCCCTAATGACAGCTGCATGTAAGCAGCCTGTACTATATCAGCCCATGCTCATTA 62446
QY 238 AACCTTTGATCTGATGCTGAC---GATGGGTGAACAGCTGCTGCTTTGATTTCAATG 294
Db 62447 AACCTTTGATCTGATGCTGAC---GATGGGTGAACAGCTGCTGCTTTGATTTCAATG 62506
QY 295 AGTCTCTAATAATACGAGCTGACCGGCTTTCGAGGCTTTGACAGAACTCTGGCTCTG 354
Db 62507 AGTCTCTAATAATACGAGCTGACCGGCTTTCGAGGCTTTGACAGAACTCTGGCTCTG 62566
QY 355 TGTTCCTCTAAGCAAGTATTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 414
Db 62567 TGTTCCTCTAAGCAAGTATTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG -GTGCA 62624
QY 415 ACAGAGCTCATCAAAAGAGAGACAGATTTTGTGACATAATATGACAGAGATTAAT 474
Db 62625 ACAGAGCTCATCAAAAGAGAGACAGATTTTGTGACATAATATGACAGAGATTAAT 62684
QY 475 ACACATATGTCATGAGAGTAATTCACATGAATAAAGTGTACAGGCTTAAGCTTTT 534
Db 62685 ACACATATGTCATGAGAGTAATTCACATGAATAAAGTGTACAGGCTTAAGCTTTT 62744
QY 535 ACGGTTAATTTCTGTCAAAAGTAAGTACAAATGCGGATCTTATCATGCTGTC---- 589
Db 62745 ACGGTTAATTTCTGTCAAAAGTAAGTACAAATGCGGATCTTATCATGCTGTCGAG 62804
QY 590 -----TCTTGAAGCCCCCTTCCCCCGCTGCCCTCCCGAGATGGGGCTGAGTCA 642
Db 62805 CATCCCTTCCGCCCCCACCCTCCACCTCCGCTCCGCTCCCGAGATGGGGCTGAGTCA 62864
QY 643 TATTTAACTGGCCATCTCAACAGTGTGTAATTAAGTGTCTTTCTTAAAGACCC 702
Db 62865 TATTTAACTGGCCATCTCAACAGTGTGTAATTAAGTGTCTTTCTTAAAGACCC 62922
QY 703 TTTCTTAACGACAAATATGCTGACCTGTACTATTAAGATCTTGTGANTATGCA----TT 757
Db 62923 TTTCTTAACGACAAATATGCTGACCTGTACTATTAAGATCTTGTGANTATGCA----TT 757
QY 758 CGGAGATTTTCTTGTGATAGTGAAGTGTGCTTCTTCACTTCCCTTACTGAG- 816
Db 62883 TTTTCTTCTTCTTGTGATAGTGAAGTGTGCTTCTTCACTTCCCTTACTGAG- 63042
QY 817 CTGACATGAGCTTCCCTGCTTCTTCTAAGTAACTGGGTAGAAATCAAGTCTCGGCTT 876
Db 63043 GCTGATGAGCTTCCCTGCTTCTTCTAAGTAACTGGGTAGAAATCAAGTCTCGGCTT 63102
QY 877 TACAGTTTAACTATTTTATGATATTCGTGAACATCATCTGTCTTCCAGATGACAA 936
Db 63103 TACAGTTTAACTATTTTATGATATTCGTGAACATCATCTGTCTTCCAGATGACAA 63162
QY 937 CTGTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 996
Db 63163 CTGTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 63219
```

RESULT 7

AC102775 185245 bp DNA linear HTG 23-MAR-2003
LOCUS Mus musculus clone RP23-115C10, WORKING DRAFT SEQUENCE, 9 unordered
pieces.
AC102775
AC102775 3 GI:29164654
HTG: HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
KEYWORDS Mus musculus (house mouse)
SOURCE
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sclerogathini; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 185245)
Birren, B., Nusbaum, C. and Lander, E.
Mus musculus, clone RP23-115C10
Unpublished
2 (bases 1 to 185245)
Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,
Anderson, S., Barna, N., Bastien, V., Boguslavsky, L., Bouckgalter, B.,
Brown, A., Camarata, J., Campolano, A., Chang, J., Chazaro, B.,
Choepe, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A.,
Cooke, P., DeArrellano, K., Dewar, K., Diaz, J. S., Dodge, S., Faro, S.,
Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S.,
Glade, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,
Hagos, B., Hearford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R.,
Jones, C., Kamat, A., Karatas, A., Kells, C., Laroque, K.,
Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Liu, G.,
Maclean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C.,
McCarthy, M., McEwan, P., McKernan, K., McPheters, R., Melrim, J.,
Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C.,
Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D.,
Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollard, V.,
Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P.,
Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupack, R.,
Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,
Topham, K., Travers, M., Travis, N., Triggillo, J., Vassiliev, H.,
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G.,
Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
Direct Submission
Submitted (23-NOV-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 185245)
Birren, B., Nusbaum, C., Lander, E., Abouelkell, A., Allen, N.,
Anderson, S., Arachchi, H. M., Barna, N., Bastien, V., Bloom, T.,
Boguslavsky, L., Bouckgalter, B., Camarata, J., Chang, J., Choepe, Y.,
Collymore, A., Cook, A., Cooke, P., Corum, B., DeArrellano, K.,
Diaz, J. S., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Faro, S.,
Ferreira, P., FitzGerald, M., Gage, D., Galagan, J., Gardyna, S.,
Graham, L., Grand-Pierre, N., Hafer, N., Hagopian, D., Hagos, B.,
Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,
Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R.,
Lindblad-Poh, K., Liu, G., Lui, A., Mabbitt, R., Maclean, C.,
Macdonald, P., Major, J., Manning, J., Matthews, C., McCarthy, M.,
Melrim, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J.,
Nguyen, C., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P.,
O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N.,
Rachupka, A., Ramsamy, U., Raymond, C., Retta, R., Rise, C., Rogov, P.,
Roman, J., Schauer, S., Schupack, R., Seaman, S., Severy, P., Smith, C.,
Spencer, B., Stange-Thomann, N., Stojanovic, N., Stubbs, M.,
Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M.,
Vassiliev, H., Venkataraman, V. S., Viel, R., Vo, A., Wilson, B., Wu, X.,
Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
Direct Submission
Submitted (23-MAR-2003) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 23, 2003 this sequence version replaced gi:2281790.
All repeats were identified using RepeatMasker:
Smit, A. F. A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: MIBR
Web site: http://www-seq.wi.mit.edu

Contact: sequence_submissions@genome.wi.mit.edu
Project Information

Center Project name: L19647
Center Clone name: 115.C.10

Summary Statistics

Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 183758 bases at least Q40
Consensus quality: 184136 bases at least Q30
Consensus quality: 184291 bases at least Q20
Insert size: 178000; agarose-fp
Insert size: 184445; sum-of-fragments
Quality coverage: 12.6 in Q20 bases; agarose-fp
Quality coverage: 12.2 in Q20 bases; sum-of-fragments

* NOTE: This is a 'working draft' sequence. It currently
* consists of 9 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 53693: contig of 53693 bp in length
53694 53793: gap of 100 bp
53794 56675: contig of 2882 bp in length
56675 56776: gap of 100 bp
56776 60589: contig of 3814 bp in length
60589 60690: gap of 100 bp
60690 67210: contig of 6521 bp in length
67210 67311: gap of 100 bp
67311 82585: contig of 15275 bp in length
82585 82686: gap of 100 bp
82686 103118: contig of 20433 bp in length
103118 103218: gap of 100 bp
103218 132145: contig of 28927 bp in length
132145 132246: gap of 100 bp
132246 182335: contig of 50090 bp in length
182335 182435: gap of 100 bp
182435 182436: contig of 2810 bp in length.

FEATURES

source

1. 185245
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/clone_1fb="RP23-115C10"
/clone_1fb="RP23-23 Female Mouse BAC"
1. 53693
/note="assembly-fragment
clone_end:SP6
vector_side:left"
misc_feature
53794..56675
/note="assembly-fragment"
56776..60589
/note="assembly-fragment"
60690..67210
/note="assembly-fragment"
67311..82585
/note="assembly-fragment"
82686..103118
/note="assembly-fragment"
103219..132145
/note="assembly-fragment"
132246..182335
/note="assembly-fragment"
182436..185245
/note="assembly-fragment
clone_end:T7
vector_side:right"

BASE COUNT 56670 a 34956 c 35069 g 57550 t 800 others
ORIGIN

Query Match 39.6%; Score 403.2; DB 2; Length 185245;
Best Local Similarity 89.7%; Pred. No. 1.9e-108;
Matches 481; Conservativity 0; Mismatches 43; Indels 12; Gaps 4;

QY	1	CCCCCTATTCCTGCTCAGATTAGAAATTCGCAAAATACCTTGACTAAAGTCGTTGG	60
DB	74	CCCCCTATTCCTGCTCAGATTAGAAATTCGCAAAATACCTTGACTAAAGTCGTTGG	133
QY	61	CTGAGAACGCTAAGACCTAAGCTGTTGAGAG---CTTGGCTTAAAGACCTGAC	117
DB	134	CTGAGAACGCTAAGACCTAAGCTGTTGAGAGCTTGGCTTAAAGACCTGAC	193
QY	118	TTCCTGAGCTCAGGGGTGACAGACCTCCGAGTCCGACAGCTGTGACACAGCAGCC	177
DB	194	TTCCTGAGCTCAGGGGTGACAGACCTCCGAGTCCGACAGCTGTGACACAGCAGCC	253
QY	178	TCCCTAATGACACGCTGCCATGTACGACCTGTAACTTACGCCATGCTCATTAAGT	237
DB	254	TCCCTAATGACACGCTGCCATGTACGACCTGTAACTTACGCCATGCTCATTAAGT	313
QY	238	AACTTGTACTGATGACGTCAC---GATGGGTGTAACGCTGCTGCTTGAATTCATAGTG	294
DB	314	AACTTGTACTGATGACGTCACGATGATGATGATGATGATGATGATGATGATGATG	373
QY	295	AGTTCTCTAAATATACAGCTGACGCTCTGCTGAGGCTTGAACAGACTGCTGCTG	354
DB	374	AGTTCTCTAAATATACAGCTGACGCTCTGCTGAGGCTTGAACAGACTGCTGCTG	433
QY	355	TGTTGCTCTACGAGATTTGTTCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTG	414
DB	434	TGTTGCTCTACGAGATTTGTTCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTG	491
QY	415	ACACGAGCTCATGAAAG---AGACAGACGATTTTACTAATTTGAAGTGAATTA	470
DB	492	ACACGAGCTCATGAAAGGAGACAGACGAGATTTTGAATTTGAAGTGAATTA	551
QY	471	ATTACACTACATTTGATGATGAGTATTTCAATGATGATGATGATGATGATGATG	526
DB	552	CTTACACTACATTTGATGATGAGTATTTCAATGATGATGATGATGATGATGATG	607

RESULT 8
AL583884
LOCUS 202342 bp DNA linear ROD 24-DEC-2002
DEFINITION Mouse DNA sequence from clone RP23-324B16 on chromosome 15,
complete sequence.
ACCESSION AL583884
VERSION AL583884.20 GI:27368255
KEYWORDS HTG.
SOURCE Mus musculus (house mouse)
ORGANISM

REFERENCE

1 (bases 1 to 202342)
TITLE Direct Submission
JOURNAL Submitted (21-DEC-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquerry@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk
On Dec 23, 2002 this sequence version replaced g1:26788018.
Sequence from the Mouse Genome Sequencing Consortium whole genome
shotgun may have been used to confirm this sequence. Sequence data
from the whole genome shotgun alone has only been used where it has
a phred quality of at least 30.

----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquerry@sanger.ac.uk

During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the

QY 303 AAAATACACCTACCGGGCTTTCAGAGGCTTTGAACAGAACTCTGCTCTGTTGGCT 362
 Db 56568 AAAAGACACCGCCCTAGCTTTTCAGAGCTTGAACAGATCTGCTGTTGGCT 56627
 QY 363 CTAAAGAGTATCTGTTCTTCTACTGCTGGGTGCTGGGTGAGTGTGTAAACAGACG 422
 Db 56628 CTATAGGAATTCCTGTTCTTCTACTGCTGGGTGCTGGGTGAGTGTGAACAGACGATG 56687
 QY 423 TCATCAAAAGAGACAGACAGTATTTTGTACATATATGAGATTAATTTACACTACA 482
 Db 56688 ATCAAAAGAGACAGACAGATTTTGTACATATATGAGATTAATTTACACTACA 56742
 QY 483 TTGTACATGAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 526
 Db 56743 TTGTACATGAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 56786
 RESULT 10
 AX163752 1022 bp DNA linear PAT 22-JUN-2001
 LOCUS AX163752 Sequence 16 from Patent WO0138579.
 DEFINITION AX163752
 ACCESSION AX163752
 VERSION AX163752.1 GI:14544858
 KEYWORDS
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1
 AUTHORS Gould-Rothberg, B.E., Dipippo, V.A., Ramseh, T.M. and Geyre, R.W.
 TITLE Method of identifying toxic agents using nsaid-induced differential
 JOURNAL gene expression in liver
 CURR PATENT: WO 0138579-A 16 31-MAY-2001;
 FEATURES
 source
 1.1022
 location/Qualifiers
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 BASE COUNT 273 a 208 c 205 g 336 t
 ORIGIN
 Query Match 28.6%; Score 290.6; DB 6; Length 1022;
 Best Local Similarity 64.5%; Pred. No. 4.5e-75;
 Matches 664; Conservative 0; Mismatches 304; Indels 61; Gaps 13;

QY 410 GTGAAACACGATCATCAAGAGAGACAGACATTTTAC--TAAATATGATAGAT 468
 Db 413 GT-----GTGATCAAAAGAGACAGATTTTACAAATATGACAAATGAG-- 460
 QY 469 TAAATTAACATCAATTTGTACATGAGTAA--TTCACTGATTAATAAGTGCAGGTAA 525
 Db 461 --ATTTCATCAATTTGTACATGAGTAAAGTGCAGGTAAAGTGCAGGTAAAGTAA 518
 QY 526 AGCTTTTAACGTTAATTTCTGTCAAAAGTATGACAAAGTGCAGGTATCATG 585
 Db 519 ATTTCGTCAAAATGACATGATATGATGAAGAGTGTGATTAATGAGAAAGTGTCT 578
 QY 586 TCTCTTTGAGCCCTCCCTCCCTGCTGCTCCCTCCCTCCCTCCCTCCCTCCCTCCCT 645
 Db 579 TAACTTTTCTCTCTCTTACCTGCAATGCTCCCTCCCTCCCTCCCTCCCTCCCTCCCT 638
 QY 646 TTAACGAGCAATATGCTGACCTGATTAATAAGTCTTGTATATGATTTGGAGAT 765
 Db 639 TTAACGAGCAATATGCTGACCTGATTAATAAGTCTTGTATATGATTTGGAGAT 698
 QY 706 TTAACGAGCAATATGCTGACCTGATTAATAAGTCTTGTATATGATTTGGAGAT 765
 Db 699 TGAAGAGCAATATGCTGACCTGATTAATAAGTCTTGTATATGATTTGGAGAT 757
 QY 766 TTTTGTGATATGATTAAGTGGCTTCTGTTTCACTCTTCACTGACCTGACCTGAC 825
 Db 758 TTTTGTGATATGATTAAGTGGCTTCTGTTTCACTCTTCACTGACCTGACCTGAC 812
 QY 826 GCTTCCCTGCTTTTCTGATTAAGTGGCTGATTAAGTGGCTGATTAAGTGGCTGAT 885
 Db 813 GCTTCCCTGCTTTTCTGATTAAGTGGCTGATTAAGTGGCTGATTAAGTGGCTGAT 872
 QY 886 TAAACATATTTAGATA-----TTGGAACATATGATTTGATTTGATTTGATTT 928
 Db 873 TAAACATATTTAGATA-----TTGGAACATATGATTTGATTTGATTTGATTTGAT 932
 QY 929 TAAACATATTTAGATA-----TTGGAACATATGATTTGATTTGATTTGATTTGAT 988
 Db 933 TAAACATATTTAGATA-----TTGGAACATATGATTTGATTTGATTTGATTTGAT 991
 QY 989 GTTCCGCTA 997
 Db 992 CTTCTCTGTA 1000
 RESULT 11
 BC050687 2302 bp mRNA linear PRI 11-APR-2003
 LOCUS BC050687
 DEFINITION Homo sapiens, ras-related C3 botulinum toxin substrate 1 (rho
 family, small GTP binding protein Rac1), clone MGC:60264
 IMAGE:6149377, mRNA, complete cds.
 ACCESSION BC050687
 VERSION BC050687.1 GI:29792301
 KEYWORDS MGC.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1
 AUTHORS Strausberg, R.
 TITLE Direct Submission
 JOURNAL Submitted (08-APR-2003) National Institutes of Health, Mammalian
 Gene Collection (MGC), Cancer Genomics Office, National Cancer
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
 USA
 NIH-MGC Project URL: <http://mgc.ncl.nih.gov>
 Contact: MGC help desk
 Email: cgabs-remail.nih.gov
 Tissue Procurement: ATCC/DCID/DTF
 CDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LIML)
 DNA sequencing by: Sequencing Group at the Stanford Human Genome

Center, Stanford University School of Medicine, Stanford, CA 94305
 Web site: <http://www-sngc.stanford.edu>
 Contact: (Dickson, Mark) mcd@paxil.stanford.edu
 Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/BLN at: <http://image.llnl.gov>
 Series: IRAC Plate: 110 Row: K Column: 14
 This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA g1: 9845510.

FEATURES

Source

Location/Qualifiers

1. 2302

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="MGC:60264 IMAGE:6149377"

/tissue_type="Skin, melanotic melanoma."

/lab_host="DH10B"

/note="Vector: PCMV-SPORT6"

189. 767

/product="ras-1 related C3 botulinum toxin substrate 1 (rho family, small GTP binding protein Rac1)"

/protein_18="AAH50687.1"

/db_xref="GI:29792302"

/db_xref="LocusID:5879"

/translation="MQAIKCVVGVGAVGKTLISYTNAPPEXITPEVDNSANYMDGPVPLGMDTADGEDYDLRLPISLPDVFICSLSPSEFVNRKMYPEVRHHCNPPIIVGKTLDRDDKDTLEKKEKLTPTITVQGLAMKEIGAVVLYESALTORCKTYFDIAIRAVLCPPVKKRKRKCLL"

BASE COUNT 617 a 524 c 498 g 663 t
 ORIGIN

Query Match 28.6%; Score 290.6; DB 9; Length 2302;
 Best Local Similarity 64.5%; Pred. No. 4.8e-75;
 Matches 664; Conservative 0; Mismatches 304; Indels 61; Gaps 13;

QY 1 CCCCTATTCTTGCAGATTAAAGAAATGCCAAATACCTTGTAAGTTGC---GTT 57
 DB 1050 CCCCATTCTTGTCAGATTAAAGAGTCCAAATACCTTGTAAGTTGCATTGTT 1109
 QY 58 GTGCGAAGACAGTAAAGCAGTAAAGTCTGTAAGTCTTGGCTTAAGAGCAGC 117
 DB 1110 GTGCGAAGACAGCAGTAAAGCAGTAAAGCAGTCTGCTTGAAGAGCAGT 1169
 QY 118 TTCTGGGCTCAGG-GGTGACAGCCCTCCCGTAGC-----TCCGACAGCCTGTGACACA 169
 DB 1170 TTCTGCACTTGAAGAGTGCAGACACTTGTCTCCATGATGATTCACAGATGGTAAAGCA 1229
 QY 170 GCACAGCCTCTTAATGACAGCCTGCCATGTACAGCAGCTGTAACTTATACGCCCATGCT 229
 DB 1230 GACAGCCTCCGATGAGAGGTGGTCCATGTAACACACAGTGAAGTACAGCAGTGT 1289
 QY 230 CATTCAGTAACTTGTACTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 289
 DB 1290 CCCCAGATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1349
 QY 290 TAGTAGTCTCTTAATAATACAGCTGACAGCAGCTTGCAGAGCTTGAACAGAACTCTGGC 349
 DB 1350 TTGTGATTTTCATACGAGTTTTCGACAGCTTTTGGGAGATTTTGAACAGAACTG-- 1407
 QY 350 TCCTGTGTGCTCTTAACGAGTATGTCTTCTAGTCTGAGGCTTGTGAGTGGTGGTGGTGGT 409
 DB 1408 ---CTATTCTCTTAATGAGAAATCTGTGTT--TAGCTGTGGGTGTGCGGGGTGGGGTGT 1461
 QY 410 GTGAAGACAGAGCTATGAAGAGACAGAGTATTTTGC--TAATATGAAGTGAAGAT 468
 DB 1462 GT-----GTATCAAGAGACAAAGACAGTATTTTGAACAAATACGAAGTGAAG-- 1509
 QY 469 TAATTACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 525

DB 1510 --ATTTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1567
 QY 526 AGCTTTTAAAGCTTAAATTTCTGTCAAAAGATGATGATGATGATGATGATGATGATGATGATGAT 585
 DB 1568 ATTTCGTCAAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1627
 QY 586 TCTCTTTGAAGCCCCCTTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 645
 DB 1628 TAACTTTTCTTCTCTCTTACACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1687
 QY 646 TTAACAGGCGCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 705
 DB 1688 TTAACAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1747
 QY 706 TTAACAGCAATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 765
 DB 1748 TGACGAGCAATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1806
 QY 766 TTTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 825
 DB 1807 TTTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1861
 QY 826 GCTTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 885
 DB 1862 GCTTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1921
 QY 886 TAACTATTTTGAATA-----TTCTGAACATCACTGCTGCTGCTGCTGCTGCTGCTGCTGCT 928
 DB 1922 TAAATATTTTGAATAATTTTAAACATGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1981
 QY 929 TACCAACATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 988
 DB 1982 TACGACATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2040
 QY 989 CTTCGGGTA 997
 DB 2041 CCTCCTGTA 2049

RESULT 12
 AK054993
 LOCUS
 DEFINITION
 Homo sapiens CDNA FLJ30431 fls, clone BRAC2008966, highly similar to RAS-RELATED C3 BOTULINUM TOXIN SUBSTRATE 1.
 ACCESSION
 AK054993
 VERSION
 AK054993.1 GI:16549633
 KEYWORDS
 oligo capping; fls (full insert sequence).
 SOURCE
 Homo sapiens (human)
 ORGANISM
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE
 AUTHORS
 1
 Tashiro, H., Yamazaki, M., Watanabe, K., Kumagai, A., Itakura, S., Fukuzumi, Y., Fujimori, Y., Komiyama, M., Sugiyama, T., Irie, R., Otsuki, T., Sato, H., Wakematsu, A., Ishii, S., Yamamoto, J., Igono, Y., Kawai, H., Saito, K., Nishikawa, T., Kimura, K., Yamashita, H., Matsuo, K., Nakamura, Y., Sekine, M., Kikuchi, H., Kanda, K., Watanabe, M., Murakawa, K., Kanehori, K., Takahashi, F., Fujii, A., Oshima, A., Sugiyama, A., Kawakami, B., Suzuki, Y., Sugano, S., Nagahara, K., Masuho, Y., Nagai, K., and Sugiyama, T.
 NEDO human cDNA sequencing project
 TITLE
 NEDO human cDNA sequencing project
 JOURNAL
 Unpublished
 2 (bases 1 to 2315)
 REFERENCE
 Isogai, T., Otsuki, T., and Sugiyama, T.
 Direct Submission
 Submitted (24-OCT-2001) Takao Isogai, Helix Research Institute, Genomics Laboratory, 1532-3 Yana, Kisarazu, Chiba 297-0812, Japan (E-mail: genomics@hri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986)
 NEDO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology (RAB); cDNA library construction: Helix Research Institute (HRI) (supported by Japan Key Technology Center etc.); 5'- 6 3'-end one pass sequencing: RAB.

HRI, and Biotechnology Center, National Institute of Technology and Evaluation; clone selection for full insert sequencing; RAB and HRI.

FEATURES
source
Location/Qualifiers
1..2315
/organism="Homo sapiens"
/mol_type="mrna"
/db_xref="taxon:9606"
/clone="BRACE2008968"
/issue_type="cerebellum"
/clone_1ib="BRACE2"
/note="cloning vector: pME18SFL3"

BASE COUNT 603 a 531 c 501 g 680 t
ORIGIN

Query Match 28.6%; Score 290.6; DB 9; Length 2315;
Best Local Similarity 64.5%; Pred. No. 4.8e-75;
Matches 664; Conservative 0; Mismatches 304; Indels 61; Gaps 13;

QY 1 CCCCTATTCCTGCTCAGATTAGAAATGGCCAAATACCTTGTGAACCTAGTGC--GTT 57
DB 1078 CCCCCTATTCCTGCTCAGATTAGAAATGGCCAAATACCTTGTGAACCTAGTGCATT 1137
QY 58 GTGCTGAGAACACGTAAAGCAGTAAAGCTGTGAGAGACTTGTCTTAAAGAGACTGCAC 117
DB 1138 GTGCCGAGAACACCGCAGCACTGAAGCTTGCAGAAAGACTTGTCTTGAAGAGCGGTAC 1197
QY 118 TTTCTGGGCTCAGG--GGTGCAGACCTCCCGTAGC-----TCCGAGCCGTGTGACAA 169
DB 1198 TTTCTCAGTTAGGAGGTGCAGACCTGTCTCTCAATGATGTTCTCAGATGCGTAAAGCA 1257
QY 170 GCACAGCCCTCTTAATGACACGCTGCATGTAAGCAGCTGTAACCTATACGCCATGCT 229
DB 1258 GAAACAGCCCTCCGAGTAAGAGCGTTGCAATGAACTCACCAGTGAAGTACACAGCTGTT 1317
QY 230 CATTACGTAATCTTACTGTACGTACAGTACAGTGGGTGAACACCTGTCTCTTGAATTCA 289
DB 1318 CCGGACATPAACATTTACTGTATGAGAGTACAGTACAGCTCTTGTGATCAGTTC 1377
QY 290 TAGTAGTCTCTTAATAATACACAGCTGACCGGCTTCCAGGCTTGAACAGAACTGTGC 349
DB 1378 TTTTGATTTTCATACGAGATTTTTCGACCAAGCTTTTCCGAGATTTTGAACAGAACTG-- 1435
QY 350 TCCCTGTGCTCTTAACGAACTATTTCTTCTAGTCTGCGGTGTGCTGGGTGAGTGT 409
DB 1436 ----CTATTCCTCTTAATGAAGAAATCTGTT--TAGCTGTGGGTGTGCCGGGTGGGTCT 1489
QY 410 GTGAACACGAGCTCATCAAGAGACAGACAGTATTTTGAC--TAATATGAAGTGAAGAT 468
DB 1490 GT-----GTGATCAAAAGACAAAGACAGTATTTTGACAAATACGAAGTGGAG-- 1537
QY 469 TAATTACACTACATTTGATGAGAGTAA--TTCACTGATTAATAAGTGCACGGGTAA 525
DB 1538 --ATTACACTACATTTGATCAAGAGATGAAGTGTACCGGTAATAAATCTTAATAAGTTA 1595
QY 526 AGCTTTTAAACGGTTAATTTCTGTCAACAGTAGATGACAAATGGCCGATCTTATCAGTG 585
DB 1596 ATTCTGTCAATGAGTAGATGAGTAAGAAAGGTGGTATATACAGGAATGTTTCT 1655
QY 586 TCTCTCTGAGCCCTTCCCTTCCCTGCTGCTCCCTCCCAAGATGGGCGGTGATCCATAT 645
DB 1656 TAAAGCTTTCCTTCTCTTACACCTGCAATGCCCTCCCAAAATGGGCATTTAATCAGCT 1715
QY 646 TTAACAGGCAATCCGACAGTTGATGTAAGTGAAGTCTTTCTTTGGACCCCTTC 705
DB 1716 TTAACAGTGTCTTCTGTTAGTCCCTAAGTGAAGTCTTTCTTATAGAACCCCTTC 1775
QY 706 TTAACAGCAATATGCTGACCTGTACTAATGAATCTTCTGATATGATTCGAGAT 765
DB 1776 TGACGAGCAATATGAGCT--CTTGTATATATAAATCTTCTGATATGATTAAGAGTT 1834
QY 766 TTTTGGGATAGTAGTAAGAGTGGCTTCCGTTTTCACCTCTTACTACGTGAGTACT 825

DB 1835 TTTTGTGATAGTAGTAAGAGTGGCTTCCATGTTAC-----TTTATTCAGAGCTAATAGT 1889
QY 826 GCTTCCCTGCTTCTTCTAGTACTGGCTGTAGAAATACAGTGTGCGGCTTACGATT 885
DB 1890 GCTTCTCTTGTAGTTCTTCTAGTACTAGGTGTAAATCAATGATGCGCTTATAGTTT 1949
QY 886 TAACTATTTTGTAGATA-----TTCTGAACATCACTGCTTGGCAGAG 928
DB 1950 TAAATATTTTGTAGTATATCTTAACACTATGAACCTTCTTACATGCTCTTGTCCAGAT 2009
QY 929 TACCAACTGTCACTGATGATGATGCGCCCTCTAGACCTCACCCAGCGGACACATG 988
DB 2010 TACCGACACTGTCACTTGCACCAATCTG--ACCTCTTACCTGCGCCAGCGGACACAG 2068
QY 989 CTTCCGGTA 997
DB 2069 CCTCTGTGA 2077

RESULT 13
HSA132695
LOCUS
DEFINITION Homo sapiens rac1 gene.
ACCESSION AJ132695
VERSION AJ132695.5 GI:8574037
KEYWORDS alternative splicing; Alu; AluJo; AluSg; AluSp; AluX; AT-rich; CT-rich; MIR; rac1 gene; rac1 protein; Rac1 protein; repetitive sequence.

SOURCE
ORGANISM Homo sapiens (human)

REFERENCE
AUTHORS Matos, P., Skaug, J., Marques, B., Beck, S., Verislimo, F., Geopach, C., Boavida, M.G., Scherer, S.W. and Jordan, P.
TITLE Small GTPase Rac1: structure, localization, and expression of the human gene

JOURNAL Biochem. Biophys. Res. Commun. 277 (3), 741-751 (2000)

MEDLINE 20517245
PUBMED 11062023
REFERENCE 2 (bases 1 to 28567)

AUTHORS Jordan, P.
TITLE Direct Submission
JOURNAL Submitted (03-FEB-1999) Jordan P., Centro de Genetica Humana, Laboratorio de Oncobiologia, Instituto Nacional de Saude, Dr. Ricardo Jorge, Avenida Padre Cruz, 1649-016 Lisboa, PORTUGAL

COMMENT On Jun 20, 2000 this sequence version replaced gi:7248282.
Related sequence AI638561.

FEATURES
Location/Qualifiers
1..28567
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="7"
/map="7p22"
/clone="H. NH0425A05, RPECI 11"

repeat_region 1..109

repeat_region 123..422

repeat_region 441..742

repeat_region 743..28117

gene 743..28117

promoter 743..1226

repeat_region 911..983

repeat_region 936

misc_feature 1026

misc_feature 1026

misc_feature 1026

misc_feature 1026

```
misc_feature /note="alternative transcription start site"
1037 /gene="rac1"
misc_feature /note="alternative transcription start site"
1053 /gene="rac1"
repeat_region /note="alternative transcription start site"
1106..1175 /rpt_family="GC-rich"
CDS join(1227..1261,13701..13772,18413..18530,22804..22860,
24268..24330,26010..26169,26458..26588)
/gene="rac1"
/note="alternative"
/codon_start=1
/product="Rac1 protein"
/protein_id="CA10733.6"
/db_xref="GI:8574039"
/db_xref="SPTRMBL:O95501"
/translacion="MQAIKCVVGDGAVGKTCLLISYTTNAPGEEYIPTVPDNSANY
MWDKPYNLGMDTAGODYDRLRPLSTPQVGHYTGKDTSRGDKPLADVFLIFS
LVSPASENVAKWPPEVRHRCPTPIIVGTKLIDLDDKDTIEKLEKLTPTTYPQ
GLAMAKELGAVKYECSALTQRLKTYFDEAIRAVLCPPEVKKRRKRCCLL"
join(1227..1261,13701..13772,18413..18530,24268..24330,
26010..26169,26458..26588)
/gene="rac1"
/codon_start=1
/product="Rac1 protein"
/protein_id="CAB53579.5"
/db_xref="GI:8574038"
/translacion="MQAIKCVVGDGAVGKTCLLISYTTNAPGEEYIPTVPDNSANY
MWDKPYNLGMDTAGODYDRLRPLSPQVDVFLICEFSLVSPASENVAKWPPEVR
HHCNPPIIVGTKLIDLDDKDTIEKLEKLTPTTYPQGLAMAKELGAVKYECSALT
TQRLKTYFDEAIRAVLCPPEVKKRRKRCCLL"
<1227..1261
/gene="rac1"
/number=1
1262..13700 /gene="rac1"
/number=1
1390..1419 /rpt_family="GC-rich"
repeat_region /rpt_family="FLAM"
1752..1793
repeat_region /rpt_family="FLAM"
1794..2005
repeat_region /rpt_family="Alu"
2140..2171
repeat_region /rpt_family="Simple_repeat"
2666..2972
repeat_region /rpt_family="Alu"
2980..3267
repeat_region /rpt_family="Alu"
3268..3289
repeat_region /rpt_family="(T)n"
3802..4102
repeat_region /rpt_family="Alu"
4123..4386
repeat_region /rpt_family="Alu"
4538..4719
repeat_region /rpt_family="LTR33A"
4754..5053
repeat_region /rpt_family="Alu"
5297..5594
repeat_region /rpt_family="Alu"
5655..5757
repeat_region /rpt_family="Alu"
5758..6062
repeat_region /rpt_family="Alu"
6063..6198
repeat_region /rpt_family="Alu"
6232..6401
repeat_region /rpt_family="Alu"
7162..7238
repeat_region /rpt_family="L1MC5"
```

```
repeat_region /rpt_family="Simple_repeat"
7348..7367
repeat_region /rpt_family="Alu"
7368..7656
repeat_region /rpt_family="Alu"
7880..8183
repeat_region /rpt_family="Alu"
8484..8778
repeat_region /rpt_family="Alu"
9071..9365
repeat_region /rpt_family="Alu"
10279..10583
repeat_region /rpt_family="Alu"
10609..10864
repeat_region /rpt_family="L2"
10970..11276
repeat_region /rpt_family="Alu"
11276..11594
repeat_region /rpt_family="Alu"
11744..11825
repeat_region /rpt_family="L2"
11764..11987
repeat_region /rpt_family="MIR"
12989..13307
repeat_region /rpt_family="Alu"
13701..13772
/gene="rac1"
/number=2
13773..18412 /gene="rac1"
/number=2
13954..14235 /rpt_family="Alu"
14807..14831
repeat_region /rpt_family="Simple_repeat"
14836..15128
repeat_region /rpt_family="Alu"
15408..15715
repeat_region /rpt_family="Alu"
15872..15896
repeat_region /rpt_family="AT-rich"
15902..16186
repeat_region /rpt_family="Alu"
16187..16314
repeat_region /rpt_family="Alu"
16413..16585
repeat_region /rpt_family="Simple_repeat"
16586..16679
repeat_region /rpt_family="FLAM_A"
16707..16810
repeat_region /rpt_family="L1M4"
16811..17108
repeat_region /rpt_family="Alu"
17109..17345
repeat_region /rpt_family="L1M4"
17377..17686
repeat_region /rpt_family="Alu"
17703..17729
repeat_region /rpt_family="Simple_repeat"
18413..18530
/gene="rac1"
/number=3
18531..22803 /gene="rac1"
/number=3
19287..19436 /rpt_family="MIR"
21161..21184
repeat_region /rpt_family="Simple_repeat"
21185..21351
repeat_region /rpt_family="Alu"
21365..21403
repeat_region /rpt_family="Alu"
21364..21365
misc_difference
```

/gene="rac1"
/note="2345 nt insertion sequence in RPI1 11-derived BAC
Query Match 28.6%; Score 290.6; DB 9; Length 28567;
Best Local Similarity 64.5%; Pred. No. 5.9e-75;
Matches 664; Conservative 0; Mismatches 304; Indels 61; Gaps 13;

QY 1 CCCCTATCTGCTGAGATTAGAAATGGCAAAATACCTGTGACACTGAGTGC---GTT 57
DB 26871 CCCCATCTCTGTTGAGATTAAAGAGTGGCAAAATACCTGTGACACTGAGTGC 26930
QY 58 GTGCTAGAGAACAGCTAGAGCTAGAGCTGTGAGAGCTTGTCTTAAAGAGCTGAGC 117
DB 26931 GTGCTAGAGAACAGCTAGAGCTAGAGCTGTGAGAGCTTGTCTTAAAGAGCTGAGC 26990
QY 118 TTCTGGGCTCAGG-GGTGGAGACCTCCGCTAGC-----TCCAGACCGGTGACACA 169
DB 26991 TTCTGGAGTAGAGAGTGGAGACACTTGTCTCTCTATGATGTTCTGAGATGCTAAAGCA 27050
QY 170 GCACAGCCTCTTAAATAGACAGCTGACATGTAAAGCAGCTGTAACTATACGCCATGCT 229
DB 27051 GAAACAGCCTCCGAAAGAGCGTTGCCATTAACCTACAGAGTAGAGACAGCTGTT 27110
QY 230 CATIAGCTACTTGTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCT 289
DB 27111 CCCGACATACATGTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCT 27170
QY 290 TAGTGAGTCTCTTAAATAGACAGCTGACAGCTGCTTGTGAGAGCTTGTGAGAGCTGAGC 349
DB 27171 TTGTGATTTCTAGAGAGCTTTTCTACACAGCTTTTGGGAGATTTTGAACAGACTG-- 27228
QY 350 TCTGTGTGCTCTTAAAGAGTATTTCTCTCTAGCTGCTGCTGCTGCTGCTGCTGCTGCT 409
DB 27229 ---CTATTCCTCTATGAGAAATCTCTGT--TACCTGGGCTGCTGCTGCTGCTGCTGCT 27282
QY 410 GTGAAGACAGAGCTATCAAGAGAGACAGAGCTATTTGAC-TAATAGAGTAGAGAT 468
DB 27283 GT-----GTGATCAAGAGACAAAGAGATTTTGAACAAATAGCAAGCTGAG-- 27330
QY 469 TAATTAACATCAATGTATGAGAGTAA---TTCACTGAATTAAGTGTCCAGGCTAA 525
DB 27331 --ATTACACTCATGTGTCAAGAGATGAAGTGTACGGTAAAGCTTAAAGAGTAA 27388
QY 526 ACCTTTTAACGTTAATTTCTGTCAACAGAGTAGATGACAAATGGCCGATCTTATCAGTG 585
DB 27389 ATTTCTGTCAAAATGAGTAGATGATGAAGAAAGGTTGATATATAGAGAAATGTTTCT 27448
QY 586 TCTCTCTGAGCCCCCTTCCCTCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 645
DB 27449 TAAGCTTTCT 27508
QY 646 TTAACGTGCGCATCTCTACAGCTGTGTAAGTGTGTAAGTGTGTAAGTGTGTAAGTGTG 705
DB 27509 TTAACGTGCTGT 27568
QY 706 TTAACGAGCAATATGTCTGACCTGTACTATTAAGATCTTCTGTATTAATGATTCGAGAT 765
DB 27569 TTAACGAGCAATATGTCTGACCTGTACTATTAAGATCTTCTGTATTAATGATTCGAGAT 27627
QY 766 TTTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 825
DB 27628 TTTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 27682
QY 826 GCTTCCCTCTGTTTCTAGTAAGTGTGTAAGTGTGTAAGTGTGTAAGTGTGTAAGTGTG 885
DB 27683 GCTTCCCTCTGTTTCTAGTAAGTGTGTAAGTGTGTAAGTGTGTAAGTGTGTAAGTGTG 27742
QY 886 TAAATATTTTAAAT-----TCTGAACATCACTGCTTGTGACAG 928
DB 27743 TAAATATTTTAAATCTTAAATCTTAAATCTTAAATCTTAAATCTTAAATCTTAAATCT 27802
QY 929 TACCAACTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 968

DB 27803 TACCAACTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 27861
QY 989 CTTCGGGTA 997
DB 27862 CCTCTGCTGA 27870

RESULT 14
AC009412 212827 bp DNA linear PRI 07-NOV-2001
LOCUS Homo sapiens BAC clone RP11-425p5 from 7, complete sequence.
DEFINITION AC009412
AC009412.6 GI:14190769
VERSION HTG.
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 212827)
Sulston, J.E. and Waterston, R.
Toward a complete human genome sequence
Genome Res. 8 (11), 1097-1108 (1998).
MEDLINE 98063792
PUBMED 9847074

REFERENCE 2 (bases 1 to 212827)
Hou, S., Maupin, R., Haekenson, W., Gregory, S. and Belter, E.
The sequence of Homo sapiens BAC clone RP11-425p5
Unpublished
REFERENCE 3 (bases 1 to 212827)
Waterston, R.H.
Direct Submision
Submitted (21-AUG-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA

REFERENCE 4 (bases 1 to 212827)
Waterston, R.H.
Direct Submision
Submitted (23-MAY-2001) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE 5 (bases 1 to 212827)
Waterston, R.
Direct Submision
Submitted (07-NOV-2001) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On May 23, 2001 this sequence version replaced gl:13431187.

COMMENT
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: sapiens@wustl.wustl.edu
----- Summary Statistics
Center project name: H_NH0425P05

NOTICE: This sequence may not represent the entire insert of this
clone. It may be shorter because we only sequence overlapping
clone sections once, or longer because we provide a small overlap
between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate
chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by sequence
from more than one subclone; and the assembly was confirmed by
restriction digest.

MAPPING INFORMATION:
The sequence of this clone was established as part of a mapping and
sequencing collaboration between the NHGRI Chromosome 7 Mapping
Project (Eric D. Green, Director), John D. McPherson in the
Department of Genetics (Washington University), and the Washington

Db 174959 GTGCCAGAGACACCGGACACTGAACTTGGCAAGACCTTCCTTTGAGAAAGCGTAGC 175018
Qy 118 TTCTGGGCTCAGG-GGTGGAGACCCGCTAGC-----TCCGAGACCGCTGACACA 169
Db 175019 TTCTGAGTTAGGAGGTGGAGACACTTGCCTCTTATGTAAGTCTCAGATGCTTAAGCA 175078
Qy 170 GCACAGCTCTCTTAATGACAGCGCTGCATGTAAGCAGCTGTAATCACTATCAGCCATGCT 229
Db 175079 GAACAGCTCTCCCAATGAAGCGTTGGCATTTGAACCTCAGCAGTAGTAGAGACCGTGT 175138
Qy 230 CATTAACATCTTTGACTGTACGTACGATGAGGCTGTACAGCTGCTCTTGTGATTCA 289
Db 175139 CCGGACATACATTTGACTGTATGATGAGGAGGAGTACAGCTAGCTCTTGTGATCAGTC 175198
Qy 290 TAGTGATCTCTAAATAATACACAGCTGACGGCTTCGACGGCTTGAACAGACTCGCC 349
Db 175199 TTGTGATTTGATGAGGAGTTTCTGACAGCTTTTGGGAGATTTTGAACAGACTG-- 175256
Qy 350 TCCTGTGTGCTCTACGAGATTTCTGCTAGTCTGAGGCTGTGCTGGTGGAGTGT 409
Db 175257 ---CTATTTCCTATGAGAAATTCGT--TACCTGTGGGTGTGCCGGTGGGTGT 175310
Qy 410 GTGAAACAGAGCTCATCAAGAGAGACAGACTATTTGAC-TAATATGAGTAGAGAT 468
Db 175311 GT-----GTGATCAAAAGGACAAAGACAGTATTTGACAAATATACAGATGGAG-- 175358
Qy 469 TAATTACATACATGTAATGAGTAA---TTCAACGTAATAAAGTGCACGGTAA 525
Db 175359 --ATTACATACATGTAATGAGTAAAGTAAAGTGTGACGGTAAACCTTAAAGGTTA 175416
Qy 526 AGCTTTTAAAGGTTAATTTCTGTCAAGAGTAGATGACAAATGGCGATCTTATCAGTG 585
Db 175417 ATTTCTGTCAAAATGAGTAGATGATGAAAGAAAGTTGGATATATCAGAAATGTTTCT 175476
Qy 586 TCTCTTTAGAGCCCTTCCCTCCCTGCTGCTCCCTCCAGATGGGGGTGAGTCAATAT 645
Db 175477 TAAGCTTTCTCTTCTTCTTACACCTGCAAGCTCCCAAAATGGCGATTTAATTCATCT 175536
Qy 646 TTAACCTGGCCATCTCACAGTGTCTAATCTTACCAAGTGGCTTTTACAGCCCTTC 705
Db 175537 TTAACCTGGTGTCTGTGTAAGTGTCTAATCTTACCAAGTGGCTTTTACAGCCCTTC 175596
Qy 706 TTAACGAGCAATATGCTGACCTGTAATTAAGATCTTCTGATTAATGCAATTCGAGATT 765
Db 175597 TGACGTAGCAATATGCTGCT--CCTGTATTAATTAATCTTCTGATTAATGCAATTCGAGATT 175655
Qy 766 TTTTGTGATATGTAAGAGTGGCTGCTGTTTTCACCTTCTTACTCAGCTGACTAGT 825
Db 175656 TTTTGTGATATGTAAGAGTGGCTGCTGTTTTCACCTTCTTACTCAGCTGACTAGT 175710
Qy 826 GCTTCCCTGCTTTCTAGTAAGTGGTGTGAGAAATCAGCTGCTGGGGCTTTACAGTTT 885
Db 175711 GCTTCCCTGCTTTCTAGTAAGTGGTGTGAGAAATCAGCTGCTGGGGCTTTTATAGTTT 175770
Qy 886 TAAACTATTTAGATA-----TCTGAAACATCACTGCTTTCAGAG 928
Db 175771 TAAATAATTTTATATATCTTAAACTATGAACTCTTAAACATCACTGCTTTCAGAGAT 175830
Qy 929 TACCAACACTGTATGATGATGATGCGCCCTCTAGACCTCAGCCAGCGGACAGATG 988
Db 175831 TACCGACACTGTACTACTGACCAATAGTG-ACCCTTCTTACTGCGCCAGCGGACAGACG 175889
Qy 989 CTTCGGTA 997
Db 175890 CCTCTGTA 175898

RESULT 15
AX408985 1232 bp DNA linear PAT 14-JUN-2002
LOCUS AX408985
DEFINITION Sequence 1632 from Patent W00229103.
ACCESSION AX408985

VERSION AX408985.1 GI:21441690
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 Alvarez, C., Horne, D., Peres-da-Silva, S. and Vockley, J.G.
Gene expression profiles in liver cancer
Patent: WO 0229103-A 1632 11-Apr-2002;
JOURNAL Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
FEATURES
source
1.1232
location/Qualifiers
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/note="EMBL/Genbank Accession No. D25274"
BASE COUNT 329 a 240 c 251 g 411 t 1 others
ORIGIN
Query Match 28 0%; Score 285.2; DB 6; Length 1232;
Best Local Similarity 64.4%; Pred. No. 1.9e-73;
Matches 658; Conservative 0; Mismatches 303; Indels 61; Gaps 13;
Qy 8 TCTTGTCAATTAAGATTTGCCAAATACCTGTGTAAGTTC---GTTGTCTGA 64
Db 1 TCTTGTCAATTAAGATTTGCCAAATACCTGTGTAAGTTC---GTTGTCTGA 60
Qy 65 GAACAGCTAAGCACTAAGCTGTTGAGAGACTTTGCTTAAAGAGCTGAGCTTGGG 124
Db 61 GAACAGCTAAGCACTAAGCTGTTGAGAGACTTTGCTTAAAGAGCTGAGCTTGGCA 120
Qy 125 CTCGAG-GGTGGAGACCTCCCGTAGC-----TCCCAACAGCTGTGACAGCAGC 176
Db 121 GTTAGAGAGGTGACACCTTCTCTTATGTAAGTTCGATGCGTAAAGAGAGCAGC 180
Qy 177 CTCTTAATGACAGCTGCTCATGTATGACGACCTGTAATATGACCATGCTCATATAG 236
Db 181 CTCCGGAATGAAGGTTGCCATTTGAACTGACAGTGTGAGGACGCTGTTCCGAGCA 240
Qy 237 TAATTTGTACTGTAGCTGACGATGGGTGTAACAGCTCTGCTTTGATTTCAATGAG 296
Db 241 TAACATTTGTACTGTATGAGTGTGAGGTGACACTGACCTGCTTGTGATGATCTTGTGA 300
Qy 297 TTTCTTAATATACAGCTGACCGCTTCTGACAGCTTGTGAAGCAAGTGGCTCGTGG 356
Db 301 TTTCTAATGACAGTTTCTGTGACAGCTTTTGGAGATTTTGAACAGACTG-----CTA 354
Qy 357 TTGCTCTTAAGAGATTTCTGCTAGTGTGAGGTGCTGGGTGAGTGTGTAAC 416
Db 355 TTTCTCTTAATGAAGATTTCTGTT--TAGCTGTGGGTGCGGGGTGGGTGTGT----- 407
Qy 417 ACAGCTCATCAAGAGAGACAGACAGTATTTTAC-TAATATGAGTACAGATTAATTA 475
Db 408 -----GTGATCAAAAGGACAAAGACAGTATTTTACAAATACAAAGTGGAG-----ATTTA 458
Qy 476 CACTACATGTACATGAGT---AATTCAGCAATAAAGTGTGACGGTAAAGCTTT 532
Db 459 CACTACATGTACAGAGATGAAGAAAGTGTGACGGGTAAAGCTTAAAGGTTAATTTCTG 518
Qy 533 TAAAGGTTAATTTCTGTCAACAGATGATGACAAATGAGCGATCTTATCAGTCTCTCT 592
Db 519 TCAATATGAGTATGATGATGAAGAAAGTGTGATATATCAGAAAGTTTTCTTAAGCTT 578
Qy 593 TGAAGCCCTTCTCCCTGCTGCTCCCTCCAGATGGGGCTTGAAGTCAATATTAAGT 652
Db 579 TTTCTTCTCTTACACTGCTGACCTCCCAATTTGGGATTTAATCATCTTTAAACT 638
Qy 653 GGCATCTCAGACATGCTTAATCTTAACTAGAGTGTCTTTCTTAAAGACCCCTCTTAAGCA 712
Db 639 GGTGTCTCTGTTAGTGTGCTTACTTATGATGAGTGTCTTTCTTAAAGACCCCTCTGACTGA 698
Qy 713 GCAATATGCTGACCTGTACTATTAAGATCTTCTGATTAATGCAATTCGAGATTTTGTGG 772

Db	699	GCAATATGCCG-CCTTGTATATAAATCTTCTGTATATGATATAGAAAGTTTTTTTCT	757
QY	773	TAGATAGTAGAAGTCGCTTCCCTGTTTTCACCTTCTTACTAGCTAGTACTGCC	832
Db	758	CGATTAGTAAAGTGCCTTCCATGTTAC-----TTTATTCAAGGCTAATAAGTGCCTTCC	812
QY	833	TTGCGTTTCTGTATACGCGGTGTAAATAATCAGTCTGTGGCGGCTTACACTTTTTTAACTA	892
Db	813	TTAGTTCCTAGTAACTAGTGTAAATAATCAGTGTGTGAGCTTTATATGTTTTTAAATAA	872
QY	893	TTTTAGATA-----TTCTGAACATCATCTGTCTTGCAGAGTACCAAC	935
Db	873	TTTTTAGATATCTTTAAACTATGAACCTTTCTTAACATCATCTGTCTTGCAGATTCACGAC	932
QY	936	ACTGTCATGTATGTATGATGCCGCCCTCTAGACTCAGCCACGCGGACATGCTTCCGG	995
Db	933	ACTGTCATGTAACCAATACGTG-ACCCTGTTTACGTGCCCGGACGCGGACACAGCCCTCTG	991
QY	996	TA 997	
Db	992	TA 993	

Search completed: August 26, 2003, 02:30:03
Job time : 3758.19 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 25, 2003, 19:52:32 ; Search time 280.632 Seconds
(without alignments)
9782.674 Million cell updates/sec

Title: US-09-717-321a-15

Perfect score: 1017
Sequence: 1 ccccatcttgctcagatt.....ccttggtgctgaggttc 1017

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_19Jun03.*

1: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*
2: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*
3: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*
4: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*
5: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*
6: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*
7: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*
8: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*
9: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*
10: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*
11: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*
12: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*
13: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*
14: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*
15: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*
16: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*
17: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*
18: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*
19: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*
20: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*
21: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*
22: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*
23: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*
24: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*
25: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1017	100.0	1017	22	AAH22397	Human rac1 gene re
2	1017	100.0	1266	22	AAH22396	Human rac1 contig
3	290.6	28.6	1022	22	AAH22398	Human rac1 gene re
4	290.6	28.6	2051	21	AAH18236	Human cancer associ
5	290.6	28.6	28567	25	ABH17030	Human MP21 gene Ra
6	285.2	28.0	1232	24	ABH83623	Human CDNA differe
7	285.2	28.0	1232	24	ABH95134	Gene #1632 used to
8	193.2	19.0	605	24	ABH38065	Human colon tumour

C	9	187.2	18.4	605	21	AAH16478
	10	172.2	16.9	432	25	ABX42447
	11	168	16.5	1318	23	AAH55569
	12	164	16.1	2167	24	ABH35339
	13	126.4	12.4	285	24	ABH71132
	14	123.2	12.1	416	25	ABH43616
	15	113	11.1	348	21	AAH22399
	16	111.6	11.0	447	21	AAH89693
	17	111.2	10.9	422	21	AAH89694
	18	103.2	10.1	3740	22	AAH07296
	19	98.2	9.7	581	24	ABH87661
	20	91.6	9.0	565	24	ABH86985
	21	87.8	8.6	506	22	AAH08004
	22	87.6	8.6	470	22	AAH18145
	23	85.6	8.4	466	22	AAH25750
	24	85.6	8.4	495	24	ABH37914
	25	68	6.7	588	21	AAH16128
	26	67	6.6	594	24	ABH60324
	27	62.2	6.1	426	25	ABH42177
	28	61.4	6.0	378	21	AAH06949
	29	59	5.8	404	25	ABH49041
	30	52.4	5.2	162	24	ABH45998
	31	41.4	4.1	901	25	ABH16468
	32	39.2	3.9	4081	23	ABH12618
	33	37.4	3.7	13712	24	ABH33531
	34	37	3.6	352	22	AAH22400
	35	37	3.6	2277	25	ABH58242
	36	37	3.6	260209	24	ABH56564
	37	36.8	3.6	3014	23	ABH06602
	38	36.8	3.6	3724	23	ABH20520
	39	36.6	3.6	995	22	AAH15320
	40	36.4	3.6	7892	24	ABH40055
	41	36	3.5	5098	22	AAH13425
	42	36	3.5	14041	22	AAH48024
	43	36	3.5	16258	24	ABH70376
	44	36	3.5	16258	24	ABH40038
	45	36	3.5	611590	21	AAH22303

ALIGNMENTS

RESULT 1
ID AAH22397 standard; DNA; 1017 BP.
XX
AC AAH22397;
XX
DT 22-AUG-2001 (first entry)
XX
DE Human rac1 gene related nucleotide sequence #1.
XX
KW Identification; toxic; hepatotoxic; differential gene expression;
KW NSAD; non-steroidal antiinflammatory drug; ds.
XX
OS Homo sapiens.
XX
PN WO200136579-A2.
XX
PD 31-MAY-2001.
XX
PF 21-NOV-2000; 2000WO-US32049.
XX
PR 22-NOV-1999; 99US-0166923.
PR 18-FEB-2000; 2000US-0183531.
PR 20-NOV-2000; 2000US-0717321.
PA (CURA-) CURAGEN CORP.
PI Gould-Rothberg BE, DiIppio VA, Ramseh TM, Getweil RW;
XX WPI; 2001-355948/37.
XX

Human colon cancer
Bovine EST associa
DNA encoding novel
Human gene express
Cott fassett-deriva
Bovine EST associa
Human rac1 gene re
Mouse Excl03 nucle
Mouse Rab2 nucleot
Human reproductive
Human colon cancer
Human colon cancer
Human breast cancer
Human breast cancer
Human breast cancer
Human colon tumour
Human colon cancer
Human cancer relat
Bovine EST associa
Human secreted pro
Bovine EST associa
CDNA encoding col
Human intracellula
Drosophila melanog
Drosophila melanog
Human immune syste
Human rac1 gene re
Human G-protein co
Human SULF2 genom
Drosophila melanog
Drosophila melanog
Human breast cance
Human breast cance
Human chemically p
Murine kit/stem ce
Internal control B
Chemically treated
Human chemically p
Arabidopsis thalia

PT Screening hepatotoxic agent comprises contacting test cell population
 PT expressing RISKMARKER or INJURYMARKER with agent, comparing expression
 PT with reference population and identifying difference in expression
 PT levels -
 XX
 PS Disclosure; Page 7-8; 76pp; English.

CC The present invention describes a method of screening a test agent for
 CC hepatotoxicity. The method comprises: (a) providing a test cell
 CC population comprising a cell capable of expressing one or more nucleic
 CC acid sequences selected from the group consisting of RISKMARKER 1-8
 CC and INJURYMARKER 1-10; (b) contacting the test cell population with a
 CC test agent; (c) measuring expression of one or more of the nucleic
 CC acid sequences in the test cell population; (d) comparing the
 CC expression of the nucleic acid sequence in the test cell population to
 CC the expression of the nucleic acid sequence in a reference cell
 CC population comprising at least one cell whose exposure status to a
 CC hepatotoxic agent is known; and (e) identifying a difference in
 CC expression levels of the RISKMARKER or INJURYMARKER sequences, if
 CC present, in the test cell population and reference cell population.
 CC The method is useful for identifying a hepatotoxic agent. The present
 CC sequence is given in the exemplification of the present invention.

SQ Sequence 1017 BP; 245 A; 245 C; 216 G; 311 T; 0 other;

Query Match 100.0%; Score 1017; DB 22; Length 1017;
 Best Local Similarity 100.0%; Pred. No. 6.9e-308;
 Matches 1017; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CCCCTATTCCTGCTCAGATTAAAGAAATGCCAAATACCTTTGGAAGTAAGTGGCTGTG 60
 1 CCCCTATTCCTGCTCAGATTAAAGAAATGCCAAATACCTTTGGAAGTAAGTGGCTGTG 60
 DB 1 CCCCTATTCCTGCTCAGATTAAAGAAATGCCAAATACCTTTGGAAGTAAGTGGCTGTG 60
 OY 61 CTGAGAACAGCTAAAGCACTAAAGCTGTGAGAGACTTGTCTTAAAGAGACTGACGCTTC 120
 61 CTGAGAACAGCTAAAGCACTAAAGCTGTGAGAGACTTGTCTTAAAGAGACTGACGCTTC 120
 DB 61 CTGAGAACAGCTAAAGCACTAAAGCTGTGAGAGACTTGTCTTAAAGAGACTGACGCTTC 120
 OY 121 TGGGCTCAGGGGCTGACAGCCCTCCGTAAGCTCCAGACCGGTGACACAGACAGCCCTCC 180
 121 TGGGCTCAGGGGCTGACAGCCCTCCGTAAGCTCCAGACCGGTGACACAGACAGCCCTCC 180
 DB 121 TGGGCTCAGGGGCTGACAGCCCTCCGTAAGCTCCAGACCGGTGACACAGACAGCCCTCC 180
 OY 181 TTAATGACAGCTCCCATGTAACGACCTGTAACCTTAATCAGCCCATGCTATTAAGTAAC 240
 181 TTAATGACAGCTCCCATGTAACGACCTGTAACCTTAATCAGCCCATGCTATTAAGTAAC 240
 DB 181 TTAATGACAGCTCCCATGTAACGACCTGTAACCTTAATCAGCCCATGCTATTAAGTAAC 240
 OY 241 TTTTACTGTAAGCTCAGATGGGTGTAACAGCTCTGCTCTTGAATTCATAGTAGTCT 300
 241 TTTTACTGTAAGCTCAGATGGGTGTAACAGCTCTGCTCTTGAATTCATAGTAGTCT 300
 DB 241 TTTTACTGTAAGCTCAGATGGGTGTAACAGCTCTGCTCTTGAATTCATAGTAGTCT 300
 OY 301 CTAAATATACGACGACGCGGCTTGACAGGCTTGAACAGAACTGGCTCTGCTGTTGC 360
 301 CTAAATATACGACGACGCGGCTTGACAGGCTTGAACAGAACTGGCTCTGCTGTTGC 360
 DB 301 CTAAATATACGACGACGCGGCTTGACAGGCTTGAACAGAACTGGCTCTGCTGTTGC 360
 OY 361 CTCTAACGAAGTATTCCTGCTAGTCGTGGGTGCTGGGTGAGAGTGTGAACAACA 420
 361 CTCTAACGAAGTATTCCTGCTAGTCGTGGGTGCTGGGTGAGAGTGTGAACAACA 420
 DB 361 CTCTAACGAAGTATTCCTGCTAGTCGTGGGTGCTGGGTGAGAGTGTGAACAACA 420
 OY 421 CGTATATCAAGGAGACAGACAGTATTTTGAATATGAGTAGAGATTAATTACACTA 480
 421 CGTATATCAAGGAGACAGACAGTATTTTGAATATGAGTAGAGATTAATTACACTA 480
 DB 421 CGTATATCAAGGAGACAGACAGTATTTTGAATATGAGTAGAGATTAATTACACTA 480
 OY 481 CATGTGATGAGATTAATTCAGTGAATTAAGTGTACGCGGTAAAGCTTTTAAAGCTT 540
 481 CATGTGATGAGATTAATTCAGTGAATTAAGTGTACGCGGTAAAGCTTTTAAAGCTT 540
 DB 481 CATGTGATGAGATTAATTCAGTGAATTAAGTGTACGCGGTAAAGCTTTTAAAGCTT 540
 OY 541 AATTTCTGTCAAAACAGTAGATGACAATGCGCGATCTTAAAGTGTCTCTTGAGCCCC 600
 541 AATTTCTGTCAAAACAGTAGATGACAATGCGCGATCTTAAAGTGTCTCTTGAGCCCC 600
 DB 541 AATTTCTGTCAAAACAGTAGATGACAATGCGCGATCTTAAAGTGTCTCTTGAGCCCC 600
 OY 601 CCTTCCCTGCTGCTCCTCCCAAGATGGGGCTGTGAGTCCATATTTAACTGGCCATCC 660
 601 CCTTCCCTGCTGCTCCTCCCAAGATGGGGCTGTGAGTCCATATTTAACTGGCCATCC 660
 DB 601 CCTTCCCTGCTGCTCCTCCCAAGATGGGGCTGTGAGTCCATATTTAACTGGCCATCC 660

OY 661 TCACAGTTGCTAACTAGCAAGTCGTTTCTTTAGACACCCCTTTAACGACCAATATG 720
 661 TCACAGTTGCTAACTAGCAAGTCGTTTCTTTAGACACCCCTTTAACGACCAATATG 720
 DB 661 TCACAGTTGCTAACTAGCAAGTCGTTTCTTTAGACACCCCTTTAACGACCAATATG 720
 OY 721 TCTGACCTGTACTAATGATCTTCTGATATGATCGGAGATTTTGGTAGATG 780
 721 TCTGACCTGTACTAATGATCTTCTGATATGATCGGAGATTTTGGTAGATG 780
 DB 721 TCTGACCTGTACTAATGATCTTCTGATATGATCGGAGATTTTGGTAGATG 780
 OY 781 AGAAGTGGTTCCTGTTTTCACCTTCCCTTACTGAGTGAAGTGCCTCTGTTT 840
 781 AGAAGTGGTTCCTGTTTTCACCTTCCCTTACTGAGTGAAGTGCCTCTGTTT 840
 DB 781 AGAAGTGGTTCCTGTTTTCACCTTCCCTTACTGAGTGAAGTGCCTCTGTTT 840
 OY 841 CTATTAAGTGGGTGATGAATATCAGCTGCTGGGCTTTTAACTATTTTATGAT 900
 841 CTATTAAGTGGGTGATGAATATCAGCTGCTGGGCTTTTAACTATTTTATGAT 900
 DB 841 CTATTAAGTGGGTGATGAATATCAGCTGCTGGGCTTTTAACTATTTTATGAT 900
 OY 901 ATTGTGAACATCACTGCTCTGCGAGATACCAACACTGTCATGTGATGCGGCCCC 960
 901 ATTGTGAACATCACTGCTCTGCGAGATACCAACACTGTCATGTGATGCGGCCCC 960
 DB 901 ATTGTGAACATCACTGCTCTGCGAGATACCAACACTGTCATGTGATGCGGCCCC 960
 OY 961 CTGTAGACCTTACCCACGCGGACACATGCTCCGTAACCTTTGGGTGTGAGGTTT 1017
 961 CTGTAGACCTTACCCACGCGGACACATGCTCCGTAACCTTTGGGTGTGAGGTTT 1017
 DB 961 CTGTAGACCTTACCCACGCGGACACATGCTCCGTAACCTTTGGGTGTGAGGTTT 1017

RESULT 2
 AAH22396/c
 ID AAH22396 standard; DNA; 1266 BP.
 XX
 AC AAH22396;
 XX
 DT 22-AUG-2001 (first entry)
 XX
 DE Human rac1 conflg SEQ ID NO:2.
 XX
 KW Identification; toxic; hepatotoxic; differential gene expression;
 KW NSAID; non-steroidal antiinflammatory drug; ds.
 XX
 OS Homo sapiens.
 XX
 PN W0200138579-A2.
 XX
 PD 31-MAY-2001.
 XX
 PE 21-NOV-2000; 2000WO-US32049.
 XX
 PF 22-NOV-1999; 99US-0166923.
 PR 18-FEB-2000; 2000US-0183531.
 PR 20-NOV-2000; 2000US-0717321.
 XX
 PA (CURA-) CURAGEN CORP.
 XX
 PI Gould-Rothberg BE, DiIppio VA, Ramseh TM, Gerweln RW;
 XX
 DR WPI; 2001-355948/37.
 XX
 PT Screening hepatotoxic agent comprises contacting test cell population
 PT expressing RISKMARKER or INJURYMARKER with agent, comparing expression
 PT with reference population and identifying difference in expression
 PT levels -
 XX
 PS Disclosure; Page 7; 76pp; English.

CC The present invention describes a method of screening a test agent for
 CC hepatotoxicity. The method comprises: (a) providing a test cell
 CC population comprising a cell capable of expressing one or more nucleic
 CC acid sequences selected from the group consisting of RISKMARKER 1-8
 CC and INJURYMARKER 1-10; (b) contacting the test cell population with a
 CC test agent; (c) measuring expression of one or more of the nucleic
 CC acid sequences in the test cell population; (d) comparing the
 CC expression of the nucleic acid sequence in the test cell population to
 CC the expression of the nucleic acid sequence in a reference cell
 CC population comprising at least one cell whose exposure status to a

CC hepatotoxic agent in known; and (e) identifying a difference in
 CC expression levels of the RISKMARKER or INJURYMARKER sequences, if
 CC present, in the test cell population and reference cell population.
 CC The method is useful for identifying a hepatotoxic agent. The present
 CC sequence is given in the exemplification of the present invention.

XX Sequence 1266 BP; 385 A; 258 C; 285 G; 338 T; 0 other;

Query Match 100.0%; Score 1017; DB 22; Length 1266;
 Best Local Similarity 100.0%; Pred. No. 7.7e-308;
 Matches 1017; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCCCTATTCCTGCTCAGATTAGAAATGGCAAAATACCTGTGCACTAGTGGCTGTG 60
 DB 1261 CCCCTATTCCTGCTCAGATTAGAAATGGCAAAATACCTGTGCACTAGTGGCTGTG 1202
 QY 61 CTGAGAACACGTTAGACACTAGCTGTGAGAGACTTGTGCTTAAAGAACATGACCTTC 120
 DB 1201 CTGAGAACACGTTAGACACTAGCTGTGAGAGACTTGTGCTTAAAGAACATGACCTTC 1142
 QY 121 TGGGGCTCAGGGGTGAGACCCCTCCGCTACTCCGACAGCGGTGAGACAGACAGCTCC 180
 DB 1141 TGGGGCTCAGGGGTGAGACCCCTCCGCTACTCCGACAGCGGTGAGACAGACAGCTCC 1082
 QY 181 TTAATGACACGGCTGCCATGTAAAGCAGCTGTAACTTATCAGCCATGCTCATTTACCTAAC 240
 DB 1081 TTAATGACACGGCTGCCATGTAAAGCAGCTGTAACTTATCAGCCATGCTCATTTACCTAAC 1022
 QY 241 TTTGTACTGTAGTCAACATGGGTGTAAACAGCTGTGCTTGTGATTTCAATGAGATTC 300
 DB 1021 TTTGTACTGTAGTCAACATGGGTGTAAACAGCTGTGCTTGTGATTTCAATGAGATTC 962
 QY 301 CTAAATATACAGCTGACCGGCTTTCGACGAGCTTTGAAAGAAAGCTGCTGGCTCCGCTGTC 360
 DB 961 CTAAATATACAGCTGACCGGCTTTCGACGAGCTTTGAAAGAAAGCTGCTGGCTCCGCTGTC 902
 QY 361 CTCTAAGCAAGTATCTGTCTAGTGTGCTGCTGGGTGAGTGTGTAACAGCA 420
 DB 901 CTCTAAGCAAGTATCTGTCTAGTGTGCTGCTGGGTGAGTGTGTAACAGCA 842
 QY 421 GCTCANTCAAGAGACAGACAGTATTTTGAATTAATGAAGTAAATTAATTTTAACTA 480
 DB 841 GCTCANTCAAGAGACAGACAGTATTTTGAATTAATGAAGTAAATTAATTTTAACTA 782
 QY 481 CATTGTACATGAGATTAATCACTGAATAAAGTGCACGGGTAAGCTTTTAAAGCTT 540
 DB 781 CATTGTACATGAGATTAATCACTGAATAAAGTGCACGGGTAAGCTTTTAAAGCTT 722
 QY 541 AATTTCTGTCAAAAGTAGATGACAAATGGCCGATCTTATCAAGTGTCTCTTGAAGCCCC 600
 DB 721 AATTTCTGTCAAAAGTAGATGACAAATGGCCGATCTTATCAAGTGTCTCTTGAAGCCCC 662
 QY 601 CCTTCCCTGCTGTCCTCCCAAGATGGGGGCTTGAGTCCATATTTAACTGCGCATCC 660
 DB 661 CCTTCCCTGCTGTCCTCCCAAGATGGGGGCTTGAGTCCATATTTAACTGCGCATCC 602
 QY 661 TCACAGTGTCTAACTAGCAAGTGTCTTCTTAAAGACCCCTTCTTAAAGCAATATG 720
 DB 601 TCACAGTGTCTAACTAGCAAGTGTCTTCTTAAAGACCCCTTCTTAAAGCAATATG 542
 QY 721 TCTGACCTGTACTATTAAGATCTTCTGATTAATGCAATGCGAGATTTTGTGTAGTAGT 780
 DB 541 TCTGACCTGTACTATTAAGATCTTCTGATTAATGCAATGCGAGATTTTGTGTAGTAGT 482
 QY 781 AGAAGTGGCTTCTGCTTTCACCTTCTTACTACAGCTAGTATGCTTCCCTGCTTTT 840
 DB 481 AGAAGTGGCTTCTGCTTTCACCTTCTTACTACAGCTAGTATGCTTCCCTGCTTTT 422
 QY 841 CTAGTACTGGGTGTGAATACAGTGTGCGGCTTAAAGTATTTTAACTATTTAGAT 900
 DB 421 CTAGTACTGGGTGTGAATACAGTGTGCGGCTTAAAGTATTTTAACTATTTAGAT 362
 QY 901 ATTCTGAACATCACTGTCTTGCCAGAGTACCAACACTGTCAATGTATGATGCGCCGCC 960

DB 361 ATTCGAAACATCACTGTCTTGCCAGAGTACCAACACTGTCAATGTATGATGCGCCGCC 302
 QY 961 CTCTAGACCTCACCACCGGGGACACATGCTTCCGCTAGCTTTGGCTGTGAGCTTC 1017
 DB 301 CTCTAGACCTCACCACCGGGGACACATGCTTCCGCTAGCTTTGGCTGTGAGCTTC 245

RESULT 3
 AAH223398
 ID AAH223398 standard; DNA; 1022 BP.

XX AAH223398;

DT 22-AUG-2001 (first entry)

DE Human rac1 gene related nucleotide sequence #2.

KW Identification; toxic; hepatotoxic; differential gene expression;

KM NSAID; non-steroidal antiinflammatory drug; ds.

XX Homo sapiens.

PN WO200138579-A2.

PD 31-MAY-2001.

PF 21-NOV-2000; 2000MO-US32049.

PR 22-NOV-1999; 9905-0166923.

PR 18-FEB-2000; 2000US-0183531.

PR 20-NOV-2000; 2000US-0717321.

PA (CURA-) CURAGEN CORP.

PI Gould-Rothberg BE, DiIippo VA, Ramseh TM, Gerwein RM;

XX WPI: 2001-355948/37.

PT Screening hepatotoxic agent comprises contacting test cell population

PT expressing RISKMARKER or INJURYMARKER with agent, comparing expression

PT with reference population and identifying difference in expression

PT levels -

XX Disclosure; Page 7-8; 76pp; English.

PS The present invention describes a method of screening a test agent for

CC hepatotoxicity. The method comprises: (a) providing a test cell

CC population comprising a cell capable of expressing one or more nucleic

CC acid sequences selected from the group consisting of RISKMARKER 1-8

CC and INJURYMARKER 1-10; (b) contacting the test cell population with a

CC test agent; (c) measuring expression of one or more of the nucleic

CC acid sequences in the test cell population; (d) comparing the

CC expression of the nucleic acid sequence in the test cell population to

CC the expression of the nucleic acid sequence in an reference cell

CC population comprising at least one cell whose exposure status to a

CC hepatotoxic agent in known; and (e) identifying a difference in

CC expression levels of the RISKMARKER or INJURYMARKER sequences, if

CC present, in the test cell population and reference cell population.

CC The method is useful for identifying a hepatotoxic agent. The present

CC sequence is given in the exemplification of the present invention.

XX Sequence 1022 BP; 273 A; 208 C; 205 G; 336 T; 0 other;

Query Match 28.6%; Score 290.6; DB 22; Length 1022;
 Best Local Similarity 64.5%; Pred. No. 3.1e-80;
 Matches 664; Conservative 0; Mismatches 304; Indels 61; Gaps 13;

QY 1 CCCCTATTCCTGCTCAGATTAGAAATGGCAAAATACCTGTGCACTAGTGGCTGTG 57
 DB 1 CCCCTATTCCTGCTCAGATTAGAAATGGCAAAATACCTGTGCACTAGTGGCTGTG 60
 QY 58 GTGCTGAGAACAGTAAAGCACTAGCTGTGAGAGACTTGTGCTTAAAGAACATGACAGC 117

```

Db      61 GTGCCGGAACACCGAGCACTGAACTTTGCAAAAGACCTTGCTTTTATGAGAGCGTAGC 120
QY      118 TTCTGGGCTCAGG-GGTGACAGACCCCTCCGTAAGC-----TCCAGACCGGTGTGACACA 169
Db      121 TTCTGCAGTTTAGAGGTGACAGACACTGCTCTCTATGTAGTTCTGAGATGCCGTAAGACA 180
QY      170 GCACAGCCCTCTTAATATACAGCGTCGCATGTAAAGCACCCTGTAACTTATACGCCCATGCT 229
Db      181 GAACAGCCCTCCGAAATAGCCGTTGGCCATTGAACTCAGCAGTGTAGAGACGCGTGT 240
QY      230 CATTACGTAATTTGTACTGTACTGTACGTACGATGGGTGTAACAGCTCTGCTCTTTGATTTC 289
Db      241 CCGGACATTAACATTGTACTGTAAATGAGTAGAGTACAGCTCAGCTCTTTGGATCAGTC 300
QY      290 TAGTGAATCTCTAATAATACAGAGTGAACCGGCTCTGACAGGCTTTGAAAGAACTGTGC 349
Db      301 TTGTGTATTCAATAGCAGATTTTCTGACACAGCTTTGCGGAGATTTTGAACAGAAAGC- 358
QY      350 TCCGTGTGTGCTCTTAACGAAGTATTTCTGCTAGTCGTGGGTGTGCTGGGTGGAGTGT 409
Db      359 ----CTATTCTCTTAATGAAGAAATTTCTGTT--TAGCTGTGGGTGTGCCGGGTGGGTGT 412
QY      410 GTGAACACGACGATCAACAAAGAGACAGACAGATTTTGTGAC-TAATATGAAGTAGAGAT 468
Db      413 GT-----GTGATCAAAAGGACAAAGACAGTATTTTGAACAAATATGCAAGTGGAG- 460
QY      469 TAATTACACTACATTTGTACATGAGTAGA--TTCAACTGAATAAAGTGTACAGGGTAA 525
Db      461 -ATTATACATACATTTGTACAGGAATGAAGAAAGTGTACAGGGTAAAACTTAAGAGTAA 518
QY      526 AGCTTTTAAAGGTTAATTTCTGTCAAAAGCATAGATACAAATGGCCGATCTTTCAGTG 585
Db      519 ATTCTGTCAATAGCATAGATGATGAAAGAAAGGTGGTGTATTCAGGAATATTTTCT 578
QY      586 TCTCTCTTGAAGCCCTCCCTCCCTGCTGTCCTCCACAGATGGGGGTGAGTGCATAT 645
Db      579 TAAGCTTTTCCCTTCTGTACACTGCACATGCCCTCCCAAAATTTGGCATTTAATTCAT 638
QY      646 TTAACATGGCCATCTCAGACAGTGTCTAATTAAGCAAGTGTCTTTCTTTAGAGACCCCTTC 705
Db      639 TTAACATGGTGTCTGTAGTCCCTAAGTAACTAGTAACTGTCTTTCTTATAGAACCCCTTC 698
QY      706 TTAACGAGCAATATGTGACCTGTACATATTAAGATCTTTTGATTAATGCAATTCGAGAT 765
Db      699 TGACTGAGCAATATGCTCCTTATTAATTAATCTTTTGATTAATGCAATTAAGAT 757
QY      766 TTTTGTGATAGTATGAGAGTGCCTGCTGTTTACCTTCTTACTACAGCTGACTAGT 825
Db      758 TTTTGTGATAGTATTAAGAGTGTTCATGCTTAC-----TTTATACAGACTTAATAGT 812
QY      826 GCTTCCCTGCTTTTCTAGTACAGTGGGTGTGAAGATCAGTGTGCGGCTTTACAGTTT 885
Db      813 GCTTCCCTGCTTTTCTAGTAACTAGGTGTAAATATGATGTGTCAGACTTTATAGTTT 872
QY      886 TAAACATTTTATAGT-----TTTGAAGCAATCAGCTGCTTCCGAGAG 928
Db      873 TAAATATTTTATAGTAAATTTCTTAACATATGAACCTTTTAACTACACTGCTTCCGAT 932
QY      929 TACCAACACTGTATGTATGATGCCGCCCTCTAGACCTCAGCCACGCGGACAGATG 988
Db      933 TACGAGACATGTCACTTGACCAATATAGT-ACCCTCTTACCTGCCGCCACGCGGACACAG 991
QY      989 CTTCCGGTA 997
Db      992 CTTCCGTGA 1000

```

RESULT 4
AAFI8236
ID AAFI8236 standard; DNA; 2051 BP.
XX
AC AAFI8236;

```

XX      14-MAR-2001 (first entry)
DT
XX      Lung cancer associated polynucleotide sequence SEQ ID 255.
DE
XX      Human; lung cancer associated protein; neuroprotective; cytosolic;
XX      cardiocastive; immunomodulatory; muscular active; vulnery;
XX      gastrointestinal; nephrotoxic; antinefctive; gynecological;
XX      antibacterial; diagnosis; neural disorder; immune disorder; reproductive;
XX      proliferative disorder; wound healing; infectious disease; ds.
XX      Homo sapient.
XX      WO20005180-A2.
XX      21-SEP-2000.
XX      08-MAR-2000; 2000WO-US05918.
XX      12-MAR-1999; 99US-0124270.
XX      (HUMA-) HUMAN GENOME SCI INC.
XX      (ROSE/) ROSEN C A.
XX      Ruben SM.
XX      WPI: 2000-587514/55.
XX      P-PSDB: AAB58360.
XX      Lung cancer associated gene sequences, referred to as lung cancer
XX      antigens, useful for treatment, prevention, and diagnosis of disorders
XX      such as lung cancer -
XX
PS      Claim 1; Page 716-717; 1425pp; English.
XX
CC      Polynucleotide sequences AAFI7982 - AAFI8424 encode human lung cancer
CC      associated proteins represented in AAB58106 - AAB58548. Lung cancer
CC      associated proteins and polynucleotide sequences, their agonists, and
CC      antagonists may have neuroprotective, cytosolic, cardiocastive;
CC      immunomodulatory; muscular active general; vulnery; gastrointestinal
CC      general; nephrotoxic; antinefctive; gynecological; or antibacterial
CC      activity. The invention also includes antibodies specific for the
CC      protein or polynucleotide sequences. The lung cancer associated
CC      polynucleotide sequences may be used for detection of lung cancer,
CC      chromosome identification, as chromosome markers, and for numerous other
CC      diagnostic or research purposes. The proteins may be used to treat
CC      disorders such as neural, immune, muscular, reproductive,
CC      gastrointestinal, pulmonary, cardiovascular, renal, and proliferative
CC      disorders. The proteins may also be used in the treatment of wounds and
CC      infectious diseases. Polynucleotide sequences AAFI8425 - AAFI8433 and
CC      peptide AAB58549 are used in the course of the invention for the
CC      identification and characterisation of the polynucleotide and protein
CC      sequences.
XX
SO      Sequence 2051 BP; 570 A; 430 C; 433 G; 612 T; 6 other;

```

Query Match 28.6%; Score 290.6; DB 21; Length 2051;
Best Local Similarity 64.5%; Pred. No. 4.5e-80;
Matches 664; Conservative 0; Mismatches 304; Indels 61; Gaps 13;

```

QY      1 CCCCTATTCTTGTGCAATTAAGATGTCGCAAAATACCTTGTGAACATTAAGTGC---GTT 57
Db      755 CCCCATTCTGTTCATTAATTAAGTGTGCAAAATACCTTGTGAACATTAAGTGTGT 814
QY      58 GTGCTGAACACGTAAGCACTAGTGTGAGAGACTTGTGCTTAAGAAAGATGACG 117
Db      815 GTGCCGAACACGAGACATGAACTTTGCAAAACCTTGTCTTTAGAAAGACGATG 874
QY      118 TTCTGGGCTCAGG-GGTGACAGACCCCTCCGTAAGC-----TCCAGACCGGTGTGACACA 169
Db      875 TTCTGAGTTAGGAGGTGACAGACACTTGTCTCTATAGTATGTTCTCAGATGCCGTAAGACA 934
QY      170 GCACAGCCCTCTTAATATACAGCGTCGCATGTAAAGCAGCTGTAACTTATCAGCCCATGCT 229

```

```

Db      | 935 GAACAGGCTCCCAATGAAGCCTTGCCATGAACTCACCAGTAGGATGACAGACGTGT 994
QY      | 230 CATTAGTACTTGTACTGTACTGTACATGCATGGGTGAACAGTCTGCTTTGATTTC 289
Db      | 995 CCCGACATTAAGTACTGTACTGTAAATGGAGTGGCGTGGACACTGACCTTTTGGATCAGTC 1054
QY      | 290 TAGTGAGTTCCTAAATATACACAGTACCGCGCTTTCAGAGCTTTTGAACAGAACTCGGC 349
Db      | 1055 TTGTGATTTCATAGGAGAGTTTCTGACAGCTTTTGGGAGATTTTGAACAGAACTG-- 1112
QY      | 350 TCCTGTGTGCTCTTAACGAGATTTCTGCTAGTCTGTGGTGTGCTGGTGGAGTGT 409
Db      | 1113 ---CTATTTCCTTAATGAAGATTTCTGTT--TAGCTGTGGTGTGCTGGGCGGTGT 1166
QY      | 410 GTGAACAGAGCTCAATGAAGAGACAGACAGTATTTTAC--TAATATAGATAAGAT 468
Db      | 1167 GT-----GTATCAAGAGACAAAGACAGTATTTTGAACAAATACAGAGTGGAG-- 1214
QY      | 469 TAATTTACACTACATTTGTACATGAGTAA--TTCAACTGAAATAAAGTGCACGGTAA 525
Db      | 1215 --ATTACACTACATTTGTACAAAGATGAAGTGTACCGGTAAAACTTAAAGTTA 1272
QY      | 526 AGCTTTTAAAGCTTAATTTCTGTCAAACAGTAGATGACAAATGGCCGATCTTATCAGTG 585
Db      | 1273 ATTTCTGTCAAAATGACAGTACATGATGAAAGAGTGTGATATATCAGAAATGTTTCT 1332
QY      | 586 TCTCTGTGAGCCGCCCTCCCTGCTGCTCCCTCCAGATGGGGCTGATGCCATAT 645
Db      | 1333 TAAGCTTTTCTCTTCTTACACCTGCGATGCTCCCAATGGCGATTAAATTCATCT 1392
QY      | 646 TTAAGTGGCCATCTCAGAGTTCCTAACTTACAGAGTGTCTTCTTCTTCTTCTTCTTCT 705
Db      | 1393 TTAAGTGGTGTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1452
QY      | 706 TTAACAGCAATATGTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 765
Db      | 1453 TGACGTGCAATATGCT--CCTGTATTAATAATCTTCTGTATTAATGCAATTAAGAGTT 1511
QY      | 766 TTTTGTGATAGTATGAGATGGCTGCTTCTTACCTTCTTCTTCTTCTTCTTCTTCTTCT 825
Db      | 1512 TTTTGTGCAATGATGAATAAGTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1566
QY      | 826 GCTTCCCTGCTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 885
Db      | 1567 GCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 1626
QY      | 886 TAACTATTTTATATA-----TTCGGAACATCACTGCTTCTTCTTCTTCTTCTTCTTCT 928
Db      | 1627 TAAATATTTTATATATTTCTTAAACTATGAACCTTCTTAAACATCACTGCTTCTTCTTCT 1686
QY      | 929 TACCAACACTGTATGATTTGATGCGCGCCCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 988
Db      | 1687 TACGACACTGTACTTGAACATATCTG--ACCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1745
QY      | 989 CTTCGGTA 997
Db      | 1746 CCTCTGTGA 1754

```

RESULT 5
ABT17030
ID ABT17030 standard; DNA: 28567 BP.

XX AC ABT17030;
XX 03-APR-2003 (first entry)
XX Human MP21 gene Rac1 SEQ ID No 4.
XX DE Cytostatic; p21 pathway modulating agent; cancer; angiogenic; apoptotic;
XX KM cell proliferation disorder; MP21; gene; ds.
XX

```

OS Homo sapiens.
XX
PN WO2003006990-A1.
XX
PD 23-JAN-2003.
XX
PF 10-JUL-2002; 2002MO-US21549.
XX
PR 12-JUL-2001; 2001US-305017P.
PR 10-OCT-2001; 2001US-328491P.
PR 15-FEB-2002; 2002US-357452P.
XX
PA (EXEL-) EXELIXIS INC.
XX
PI Friedman L, Plozman GD, Belvin M, Li D, Funke RP;
DR WPI; 2003-221779/21.
DR P-PDB; ABJ19756.
XX
PT Identifying candidate p21 pathway modulator, by contacting an assay
PT system having modifiers of p21 polypeptide or gene with a test agent to
PT provide a reference activity in system and detecting test agent-biased
PT activity.
XX
PS Examples; Page 56-72; 199pp; English.
XX
CC The invention relates to a novel method for identifying a candidate p21
CC pathway modulating agent. The novel method comprises contacting an assay
CC system, comprising a purified MP21 polypeptide (modifier of p21) or
CC nucleic acid, with a test agent under conditions, so that but for the
CC presence of a test agent, the assay system provides a reference activity
CC and detection of test agent-biased activity of the assay system. The
CC novel method of the invention is useful for identifying a candidate p21
CC pathway modulating agent. The invention also includes a method for
CC modulating the p21 pathway of a cell, and a method for diagnosing a
CC disease e.g. cancer in a patient. The identified modulators are useful in
CC diagnosis, therapy and pharmaceutical development. The modulators are
CC useful in a variety of diagnostic and therapeutic applications including
CC angiogenic, apoptotic and cell proliferation disorders. This
CC polynucleotide sequence represents a gene encoding an MP21 protein of the
CC invention.
XX
SQ Sequence 28567 BP; 6762 A; 6358 C; 6796 G; 8651 T; 0 other;

```

Query Match 28.6%; Score 290.6; DB 25; Length 28567;
Best Local Similarity 64.5%; Pred. No. 2e-79;
Matches 664; Conservative 0; Mismatches 304; Indels 61; Gaps 13;

```

QY 1 CCCCTATTTCTTGTCTCAGATTAAAGATTTGCCAAATACCTTGTGAACTAAGTTGC--GTT 57
Db 26871 CCCCATTTCTTGTCTCAGATTAAAGATTTGCCAAATACCTTGTGAACTAAGTTGTT 26930
QY 58 GTGCTGAGAACAGCTAAGCACTAAGCTGTGTGAAGACTTGTCTTAAAGAACTGCAGC 117
Db 26931 GTGCCGAGAACAGCGAGCACTTGTGCAAGACCTTGTCTTGTGAAGAAAGCGTAGC 26990
QY 118 TTCTGGGCTCAGG--GGTGCAGACCTTCCCTAGC-----TCCAGACCGTGTGACACA 169
Db 26991 TTCTGCAGTAGGAGGTGGCAGACACTTGTCTCTCTATGTATGTCTCAGATGCTTAAGCA 27050
QY 170 GCAAGCTCTCTTAATGACACGCTGCCATGAAGCAAGCACTGTAACTTATCAGCCATGCT 229
Db 27051 GAACAGCTCTCCGATGAAGAGCGTTGCCATTTGAACCTACACAGTAGTAGACAGCTGTT 27110
QY 230 CATTAGTAACTTTGTACTGTACTGTACATGCATGGGTGAACAGTCTGCTTTGATTTC 289
Db 27111 CCCGACATTAAGTACTGTACTGTAAATGGAGTGGCGTGGACACTGACCTTTTGGATCAGTC 299
QY 290 TAGTGAGTTCCTAAATATACACAGTACCGCGCTTTCAGAGCTTTTGAACAGAACTCGGC 349
Db 27171 TTGTGATTTCATAGGAGAGTTTCTGACAGCTTTTGGGAGATTTTGAACAGAACTG-- 27228
QY 350 TCCTGTGTGCTCTTAACGAGATTTCTGCTAGTCTGTGGTGTGCTGGGCGGTGAGTGT 409

```

```

Db 27229 ----CTATTTCCTTAATGAAGAAATTCGTT--TAGCTGGGGTGTGCGGGGTGGGTGT 27282
QY 410 GTGAACACGACGATCATCAAGAGACAGACAGATATTTGAC-TAATATGATAGAGAT 468
Db 27283 GT-----GTGATCAAAAGGACAAAGACAGATATTTTGCAAAATATGAGAGTGAG-- 27330
QY 469 TAAATTAACACTACATTTGACATGAGATGA---TTCAACTGAATAAAGTGTACGGGTGA 525
Db 27331 --ATTATGACATACATTTGACAAAGAAATGATGACGGGTGATAAAACTTAAGAGTTA 27388
QY 526 AGCTTTTAAAGGTTAATTTCTGTGCAAAACATAGATGCAATAGCCGATCTTATCACTG 585
Db 27389 ATTCTGTCAATGCAATGATAGATGAAGAAAGAGTGGTATTTATCAGGAATGTTTCT 27448
QY 586 TCTCTCTGAGCCCTCCCTCCCTGCTGCTGCCAGATGGGGGTGAGTCATAT 645
Db 27449 TAAGCTTTTCTTCTTCTTACACCGCATGCCATGCCCAAAATTTGGGCAATTAAATTCATCT 27508
QY 646 TTAACATGCGCATCCTCACAGTGTCTTAACATTAGACAGTGTCTTCTTATGACCCCTTC 705
Db 27509 TTAACATGCTGTGTCTGTATGCTTACCTTAAGTGAAGTCTTTCTTATGACACCCCTTC 27568
QY 706 TTAAGAGCAATATGCTGACCTGACTATGAATCTTTGATTAATGCAATTCGGAGATT 765
Db 27569 TGACTGAGCAATATGCTCCTTGTATTTATTAATCTTTCTGATATGCAATTAGAGATT 27627
QY 766 TTTTGTGATAGTATGAGAGTGCCTGCTGTTTTCACCTCTCTTACTACAGCTGACTAGT 825
Db 27628 TTTTGTGATAGTATGAGAGTGCCTGCTGTTTTCACCTCTCTTACTACAGCTGACTAGT 27682
QY 826 GCTTCCCTGCTGTTTCTGATACGCGGTGAGAAATCAAGCGTCCGCGCTTACAGTCTT 885
Db 27683 GCTTCCCTGCTGTTTCTGATACGCGGTGAGAAATCAAGCGTCCGCGCTTACAGTCTT 27742
QY 886 TAAATATTTAGATTA-----TTCTGAACATCACTGTCTTCCAGAG 928
Db 27743 TAAATATTTAGATTAATTTCTTAACATGTAACCTTCTTATCATGACGTCTTCCAGAT 27802
QY 929 TACCAACACTGTATGATGATGCGCCCTCTAGACCTACACCGCGGACATG 988
Db 27803 TACCGACACTGTCACTGACCAATATGACGCTCTTACTGCGCCACGCGGACACACG 27861
QY 989 CTTCGGGTA 997
Db 27862 CTTCCTGTA 27870

```

RESULT 6
ABR83623
ID ABR83623 standard; cDNA: 1232 BP.

AC ABR83623;

XX 14-AUG-2002 (first entry)

XX Human cDNA differentially expressed in granulocytic cells #194.

XX Human; ss; granulocytic cell; DNA chip; bacterial infection;
XX viral infection; parasitic infection; protozoal infection;
XX fungal infection; sterile inflammatory disease; psoriasis;
XX rheumatoid arthritis; glomerulonephritis; asthma; thrombosis;
XX cardiac reperfusion injury; renal reperfusion injury; ARDS;
XX adult respiratory distress syndrome; inflammatory bowel disease;
XX Crohn's disease; ulcerative colitis; periodontal disease;
XX granulocyte activation; chronic inflammation; allergy.

OS Homo sapiens.

PN WO200228999-A2.

XX 11-APR-2002.

XX

```

PF 03-OCT-2001; 2001WO-US30821.
XX
PR 03-OCT-2000; 2000US-237189P.
PA (GENE-) GENE LOGIC INC.
XX
PI Beazer-Barclay Y, Weissman SM, Yamaga S, Vockley J;
XX WPI; 2002-435328/46.
DR
XX
PT Detecting granulocyte activation by detecting differential expression
PT of genes associated with granulocyte activation, which serves as
PT diagnostic markers that is useful for monitoring disease states and
PT drug toxicity
XX
PS Claim 1; SEQ ID No 194; 114pp; English.
XX
CC The invention relates to detecting (M1) granulocyte (GC) activation
CC (GCA), by detecting the level of expression of gene(s) (Gs) identified by
CC DNA chip analysis as given in the specification, and comparing
CC the expression level to an expression level in an unactivated
CC GC, where differential expression of Gs is indicative of GCA.
CC Also included are modulating (M2) GA by contacting GC with an agent
CC that alters the expression of at least one gene in Gs; (2) screening (M3)
CC for an agent capable of modulating GCA or an inflammation (especially
CC chronic) in a tissue, an allergic response in a subject, exposure of a
CC subject to a pathogen or sterile inflammatory disease using the
CC gene expression profile; (3) detecting (M4) an inflammation (especially
CC chronic) in a tissue, an allergic response in a subject, exposure of a
CC subject to a pathogen or sterile inflammatory disease, by detecting the
CC level of expression of a sample of the tissue of gene(s) from Gs, where
CC the level of expression of the gene is indicative of inflammation;
CC (4) treating (M5) an inflammation (especially chronic) or in a tissue,
CC an allergic response in a subject, exposure of a subject to a pathogen
CC or sterile inflammatory disease, by contacting a tissue having
CC inflammation with an agent that modulates the expression of gene(s)
CC from Gs in the tissue. M1 is useful for detecting GCA; M2 is useful for
CC modulating Gs; M3 is useful for screening an agent capable of modulating
CC GCA preferably in an inflammation in a tissue; M4 is useful for
CC detecting an inflammation (especially chronic) in a tissue, an allergic
CC response in a subject, exposure of a subject to a pathogen or sterile
CC inflammatory disease (e.g. psoriasis, rheumatoid arthritis,
CC glomerulonephritis, asthma, thrombosis, cardiac reperfusion injury, renal
CC reperfusion injury, ARDS, adult respiratory distress syndrome,
CC inflammatory bowel disease, Crohn's disease, ulcerative colitis,
CC periodontal disease; also bacterial infection, viral infection,
CC parasitic infection, protozoal infection, fungal infection and M5 is
CC useful for treating one of the above conditions. The present
CC sequence represents a gene differentially expressed in granulocytes.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1232 BP; 329 A; 240 C; 251 G; 411 T; 1 other;

```

Query Match 28.0%; Score 285.2; DB 24; Length 1232;
Best Local Similarity 64.4%; Pred. No. 1.7e-78;
Matches 658; Conservative 0; Mismatches 303; Indels 61; Gaps 13;

```

QY 8 TCTTGCTGAGATTAAAGTATGCAAAATATACCTGTGACATTAAGTTGC---GTTGTCGGA 64
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1 TCTTGTATGATTAAGATTTGCAAAATATACCTTCTGAACTACACTGATTTGTGGCGGA 60
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 65 GAACACGTATGACACTTACCTGTGAGAGACTTGTCTTAAAGAACTGACACTTCTGGG 124
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 61 GAACACGAGACAGTGAACCTTGTGCAAAAGACCTTGTGAGAGAGCTGATTCGCA 120
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 125 CTCAGG-GGTGACAGCCTCCCGTAGC-----TCCAGACCGGTGACACAGCAGC 176
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 121 GTTAGAGAGTGCAGACACTTCTCTCTATGTAGTTCTCAGATGCGTAACAGAAACAGC 180
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 177 CTCCTTAATGACAGCCTCCATGTAAAGCACCACTGTAACTTAATACAGCCCATGCTCATACG 236

```


KW detect mutation: non-invasive diagnostic method; ds.
 XX Homo sapiens.
 OS
 XX MO200012702-A2.
 PN
 XX 09-MAR-2000.
 PD
 XX 30-AUG-1999; 99MO-US19424.
 PF
 XX 31-AUG-1998; 98US-0098639.
 PR 27-JAN-1999; 990S-0117393.
 XX
 XX (FARB) BAYER CORP.
 PA
 PI Endege WO, Steinhann KE, Astle JH, Burgess CC, Carroll E;
 PI Catino TJ, Dwyer P, Ford DM, Lewis ME, Molino GA, Monahan JE;
 PI Schlegel R;
 XX
 XX WPI; 2000-256641/22.
 DR
 XX
 XX Novel nucleic acids and proteins for identifying therapeutic agents
 PT useful for treating and diagnosing cancer, especially colon cancer -
 PT
 XX
 PS Claim 16; Page 319; 345pp; English.
 XX
 XX This sequence represents a human nucleotide sequence which is
 CC differentially expressed in colon cancer cells compared to the expression
 CC levels in normal cells. The nucleotide sequence can be used as a source
 CC of primers and probes. The nucleotide sequence is useful for determining
 CC the phenotype of a cell by detecting the differential expression of the
 CC sequence relative to a normal cell. The probes derived from the sequence
 CC can also be used to determine the phenotype of cells in a sample. Probes
 CC and antibodies which hybridise to the nucleotide sequence can also be
 CC used to determine the phenotype of a cell. The primers are useful for
 CC detecting a mutation in a test nucleotide sequence and also for detecting
 CC cancer, preferably colon cancer. Antibodies against the protein encoded
 CC by the nucleotide sequence can also be used in a method to detect colon
 CC cancer. The diagnostic method is non-invasive and accurate for diagnosing
 CC colon cancer at an early stage.
 XX
 SQ Sequence 605 BP; 214 A; 105 C; 116 G; 155 T; 15 other;
 Query Match 18.4%; Score 187.2; DB 21; Length 605;
 Best Local Similarity 65.5%; Pred. No. 5.9e-48;
 Matches 348; Conservative 0; Mismatches 158; Indels 25; Gaps 5;
 QY 485 GTACATGAGTATTCACCTGATTAATGAAGTGCAGGTTAAAGCTTTTAAACGGTTAAT 544
 DB 597 GCACGANGANNAAAGTGCNGGGTAAACCTTAAAGGTTAATTTTGTCAATNCAGTA 538
 QY 545 TCTGCTAACAGTAG-ATGACAAATGGCGGATCTTATCAGTGTCTCTGAGCCCCCT 603
 DB 537 GATTAANNAANNAAGGTTTNTATACACGANAATGTTTCTTANGGTTTCTTNTTCT 478
 QY 604 TCCCTCTGCTGCCCTCCCAAGATGGGGCTTGAGTCATATTAACTGGCATCTCA 663
 DB 477 AACACTGGCATGCGNCCCAATGGGATTTAATCATCTTTAAACAGCTTGTCTCT 418
 QY 664 CAGTGTCTACTAGCACTGCTTTCTTTAGACCCCTTCTTAAGACATATATGCT 723
 DB 417 TAGTGCTAACTAGTAAGTCTTCTTCTTAAGAACCCCTCTGACTGAGCAATATGCT 358
 QY 724 GACCTGACTATTAAGTCTTCTGANTATGCAATTCGATGATTTTGGTAGATATTA 783
 DB 357 -CCTGTATTAATTAATTAATCTTCTGATTAATGCAATTAAGGTTTGTGATTAATA 299
 QY 784 AGTGGCTCTGTTTCACTTCTTCTTACTACAGCTAGTACTGCTCCCTGCTTTCTA 843
 DB 298 AGTGGCTTCCATGTTAC-----TTTATTCAGAGCTAATAAGTCTTCTTCTTCTTA 244
 QY 844 GTAACTGGGTGTAAGAAATACGCTGCGGCTTTACAGTTTAACTATTTAGATA-- 901
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

DB 243 GTAACTAGGTGTAAAAATCATGTGTGACGCTTATATAGTTTAAATATTTAGATAAT 184
 QY 902 -----TTCTGAACATCAGTGTCTTCCAGAGACCAACATGTCATG 946
 DB 183 TCTTAACATATGACCTTCTTAAATCATCAGTGTCTTCCAGATTAACCGACATGCTCACTTG 124
 QY 947 ATGTATCGCGCCCTCTAGACCTCAACGCGGACATACCTGCTCCGTA 997
 DB 123 ACCAATATCTG-ACCTCTTACTCTGCCCGGACGCGGACACACGCTCTCTGTA 74
 RESULT 10
 ID ABX42447
 ABX42447 standard; cDNA; 432 BP.
 AC ABX42447;
 XX
 XX 20-FEB-2003 (first entry)
 DE
 XX Bovine EST associated with lactation/muscle/fat deposition #7612.
 DE
 XX
 KW Bovine; ss; EST; expressed sequence tag; lactation; LMPD;
 KW muscle deposition; fat deposition; genome mapping; gene identification;
 KW gene analysis; cattle breeding.
 XX
 OS Bos Taurus.
 XX
 PN US2002137139-A1.
 XX
 XX 26-SEP-2002.
 PD
 XX 24-SEP-2001; 2001US-0960352.
 PF
 XX 12-JAN-1999; 99US-115707P.
 PR 11-JAN-2000; 2000US-0480902.
 XX
 XX (BYAT/) BYATT J C.
 PA (MATH/) MATHIALAGAN N.
 PA (TAON/) TAO N.
 PA (WAR/) WARREN W C.
 XX
 XX Byatt JC, Mathialagan N, Tao N, Warren WC;
 PI
 DR WPI; 2003-110599/10.
 XX
 PT New nucleic acid associated with lactation, and muscle and fat
 PT deposition, useful for genome mapping, gene identification and
 PT analysis, cattle breeding, or for genetically improving cattle -
 PS
 PS Claim 2; SEQ ID No 7612; 245pp; English.
 XX
 CC The invention relates to a purified nucleic acid molecule associated with
 CC lactation or muscle and fat deposition (designated LMPD), derived
 CC from cattle, and the LMPD nucleic acid can specifically hybridise to a
 CC second nucleic acid molecule comprising any of 15112 nucleotide
 CC sequences, appearing as ABX34836-ABX49947, or complements of them.
 CC Also included are: (1) a transformed cell having a nucleic acid
 CC comprising an LMPD nucleic acid linked to a promoter and a 3' non-
 CC translated sequence that functions in the cell to cause termination of
 CC transcription and addition of polyadenylated ribonucleotides to a 3' end
 CC of the mRNA molecule; and (2) determining a level or pattern of a
 CC molecule in a bovine cell or tissue comprising: (a) incubating a marker
 CC nucleic acid (comprising any of the 15112 nucleic acid sequences or its
 CC complement or fragment) with a complementary nucleic acid molecule
 CC obtained from the bovine cell or tissue, where hybridisation between the
 CC marker nucleic acid and the complementary nucleic acid permits the
 CC detection of the molecule; and (b) detecting the level or pattern of the
 CC complementary nucleic acid, where the detection of the complementary
 CC nucleic acid is predictive of the level or pattern of the molecule.
 CC The LMPD nucleic acid is used for determining a level or pattern
 CC of a molecule in a bovine cell or tissue. It is useful for genome
 CC mapping, gene identification and analysis, cattle breeding, preparation
 CC of constructs for use in cattle gene expression, or for genetically

CC Improving cattle. The present sequence is one of the 15112 bovine
CC LMD EST (expressed sequence tag) nucleic acids.
CC Note: The present sequence was not shown in the specification but
CC was obtained in electronic format from the USPTO web site:
CC seqdata.uspto.gov/sequence.html?docid=20020137139.

XX
SQ Sequence 432 BP; 123 A; 76 C; 78 G; 155 T; 0 other;

Query Match 16.9%; Score 172.2; DB 25; Length 432;
Best Local Similarity 72.2%; Pred. No. 2,4e-43;
Matches 294; Conservative 0; Mismatches 103; Indels 10; Gaps 5;

QY 486 TACATGGAGTATCACTGATTAAGTGCAGCGGTAAGCTTTTACGGTTAATT 545
DB 27 TACACGTAAACATTAAACTATTTAAAGCATACGGGTAAGGTTTAAAGGTTAATT 86
QY 546 CTGCACAA--CAGTAGATAGACAATGGCCGATCTTATCAGTCTCTTGAAGCC-- 600
DB 87 CTGCAATATGGGATGATGAGAAAGGCGGATATACAGTGTCTTTTAAAGCTTT 146
QY 601 -CCTTCCCTGCTGCTCCCTCCAGATGGGCGTGAATCCATATTTAAACGCGCATC 659
DB 147 TCCCTAAACCTGCGCCACCTCTGAAATTTGGCATTTAAATCATCTGACCTGTCATC 206
QY 660 CTCACAGTGTACTATAGCAAGTCTTTCTTAGAGCCCTCTTAAAGCAATAT 719
DB 207 CTCATAGTGTACTATAGCAAGTCTTTCTTAGAGCCCTCTTAAAGCAATAT 266
QY 720 GTCGACCTGTACTATAGATCTTTCGATATATGCAATTCGAGATTTTGTAGATAG 779
DB 267 GCCT-CCTTGTATATATAATCTTTCGATATATGCAATTTTGTAGATAG 325
QY 780 TAGAAGGCGTCTCTGTTTTCACCTTCTTACTC--AGTCACTAGTGTCTCTGCT 837
DB 326 TAAAGGCGACCTCTGTTTTCATGTGTCTTATTCAGCAAGCTAATAGCTTCTTACT 385
QY 838 TTTCTAGTAACT-GGGTGTAGAAATCAAGTCTGCGCTTACAGTT 883
DB 386 TTTCTAGTAACTAGTATATAAATTCACGTTGTGACGCTTATGCTT 432

RESULT 11
AAS65569
ID AAS65569 standard; cDNA: 1318 BP.

XX AAS65569;

XX 13-FEB-2002 (first entry)

XX DNA encoding novel human diagnostic protein #1373.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder; ss.

XX Homo sapiens.

XX WO2001.75067-A2.

XX 11-OCT-2001.

XX 30-MAR-2001; 2001WO-US08631.

XX 31-MAR-2000; 2000US-0540217.

XX 23-AUG-2000; 2000US-0649167.

XX (HXSE-) HXSEQ INC.

XX Drmanac RT, Liu C, Tang YT;

XX WPI: 2001-639362/73.

XX P-PSDB; ABG01382.

XX New isolated polynucleotide and encoded polypeptides, useful in

PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity

XX Claim 1; SEQ ID No 1373; 103bp; English.

XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful for medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 1318 BP; 346 A; 268 C; 279 G; 424 T; 1 other;

Query Match 16.5%; Score 168; DB 23; Length 1318;
Best Local Similarity 58.1%; Pred. No. 9.4e-42;
Matches 615; Conservative 0; Mismatches 365; Indels 78; Gaps 15;

QY 1 CCCCTATCTTGTGCAGATTAAGATGCCAAATACCTGTGAACTAGTTGC---GTT 57
DB 39 CCCCATTCTTGTGCAGATTAAGATGCCAAATACCTGTGAACTAGATGATGTT 98
QY 58 GTGCTGAGACACGTAAGCACTTAAGCTGTGAGAGCTTTGTC-CCTTAAAGACTGAC 116
DB 99 GTGCGAGAACACCGAGCACTGAACCTTTCGATTTGAGAGACGGTAG 158
QY 117 CTTGCGGCTCAGG-GGTGACAGCCCTCCGTAGC-----TCCGACAGCGGTGACA 167
DB 159 CTTCTGCAGTTAGAGAGGTGACAGACCTTCTCTCATGTAGTTCTCAGATGCGTAAG 218
QY 168 CAGCAGACCTCTTAATGACACGCTGCATGTAAGCAGCTTAATCTTA-TCAAGCCAT 226
DB 219 CAGAAAGCCCTCCGAATGAAAGCGTGCATGTAAGTCAACAGAGATGACGACGCT 278
QY 227 GCTCATTAAGTAACTTTGTACTGTACGTCACGATGGGTGTACAGCTCTGCTTTGATT 286
DB 279 GTTCCCGCATTAATGTTACTGTATGGA--GTGAGGCTGACAGCTCAACCTCTTGA 336
QY 287 TCATAGTAGTTCTCTAAATATACAGCTGACGCGCTTGCAGGCTTTGAACAGACTCT 346
DB 337 TCAGTCTTTGTGATTTCAATACGAGTTTCTGACACAGCTTTTGGGGAGATTTGAACAG 396
QY 347 GGTCTCTGTGCTCTTAACGAAGTATCTGTTCTAGTGTGGGTGCTGGGTGAG 406
DB 397 AACTGGCTATTTCTCTTAATGAGAAATTTGTTAGCTGTGGGGGTGCTGGGTGAG 456
QY 407 TGTGTGAAGACAGAGCTCATCAAGAGACAGACAGTATTTTGTGACTAATATGAAGTACAG 466
DB 457 GGTGTGTGTGTG---ATCAAAAGACAAAGACAGTAAATTT---TGGAAACAAATATCGG 508
QY 467 ATTAAATTACACTATGATGATGAGTAAATTAATCAAGTAATTAAGTGCAGGGTAA 526
DB 509 AGTGGAGAGATTTACCACTACCACTTTGTTACAGGAAATGAAGTGTACGCGGTAA 568
QY 527 GCTTTTAAACGTTAATTTCTGTCAAAACAGTATGACAAATGCGATCTTATACAGTGT 586
DB 569 CTCTTAAAGGTTAATTTCTGTGTCATGAGTATGATGAAAGAAAGGTTGGTAT 628

QY 676 TACCAAGTCCTTCTTTCAGACCCCTTCTTAACGCAATATGTCGACTACTAT 735
DB 416 TAGTAAAGTGTCTTCTTCTTAAGAACTATCTTATAGCAATATGCTC-CTGTATAT 358
QY 736 AAGATCTTCTTGTATATGATTCGAGATTTTTCGTAGATAGTAAAGTGGTCTG 795
DB 357 AAAATCTTCTGATATGATTAAGAAATTTTTCGTAGCTTATGAAAGTGCACCTG 298
QY 796 TTTTCACCTTCCTTCTTCTC--AGCTGACATGCTGCTTCTGTTTCTAGTACT- 852
DB 297 TTTTCAATGTTGCTTATTCAAAGCTAATTAAGTCTTCTTCTAGTCTGATAGGT 238
QY 853 TGTAGAAATCAGCTGCTGCGCTTACAGTTTAACTATTTAGATATTCGAAC 910
DB 237 TATTAATAATCAGCTGTTGACGCTTATGTTTAAAGATCTCAGATATTCCTCAAC 180

RESULT 15
AAH2399
ID AAH2399 standard; DNA; 348 BP.
XX
AC AAH2399;
XX
DT 22-AUG-2001 (first entry)
XX
DE Human rac1 gene related nucleotide sequence #3.
XX
KM Identification: toxic; hepatotoxic; differential gene expression;
KM NSAID; non-steroidal antiinflammatory drug; ds.
XX
OS Homo sapiens.
XX
PN MO200136579-A2.
XX
PD 31-MAY-2001.
XX
PF 21-NOV-2000; 200OWO-0532049.
XX
PR 22-NOV-1999; 99US-0166923.
PR 18-FEB-2000; 200OUS-0183531.
PR 20-NOV-2000; 200OUS-0717321.
XX
PA (CURA-) CURAGEN CORP.
XX
PI Gould-Rothberg BE, DiPippo VA, Ramsehl TM, Gerwehn RW;
XX
DR WPI; 2001-355948/37.
XX
PT Screening hepatotoxic agent comprises contacting test cell population
PT expressing RISKMARKER or INJURYMARKER with agent, comparing expression
PT with reference population and identifying difference in expression
PT levels -
XX
PS Disclosure; Page 8-9; 76pp; English.
XX
CC The present invention describes a method of screening a test agent for
CC hepatotoxicity. The method comprises: (a) providing a test cell
CC population comprising a cell capable of expressing one or more nucleic
CC acid sequences selected from the group consisting of RISKMARKER 1-8
CC and INJURYMARKER 1-10; (b) contacting the test cell population with a
CC test agent; (c) measuring expression of one or more of the nucleic
CC acid sequences in the test cell population; (d) comparing the
CC expression of the nucleic acid sequence in the test cell population to
CC the expression of the nucleic acid sequence in a reference cell
CC population comprising at least one cell whose exposure status to a
CC hepatotoxic agent is known; and (e) identifying a difference in
CC expression levels of the RISKMARKER or INJURYMARKER sequences, if
CC present, in the test cell population and reference cell population.
CC The method is useful for identifying a hepatotoxic agent. The present
CC sequence is given in the exemplification of the present invention.
XX
SQ Sequence 348 BP; 103 A; 76 C; 67 G; 102 T; 0 other;

Query Match 11.1%; Score 113; DB 22; Length 348;
Best Local Similarity 100.0%; Pred. No. 7.7e-25;
Matches 113; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 905 TGAACATWCACCTGTCTTGCAGAGTACCAACTGTCATGTGATGAGCCGCCCTCT 964
DB 1 TGAACATWCACCTGTCTTGCAGAGTACCAACTGTCATGTGATGAGCCGCCCTCT 60
QY 965 AGACCTCACCCACGCGGACACATGCTCCGCTACCTTTGGGCTGTGAGGTTC 1017
DB 61 AGACCTCACCCACGCGGACACATGCTCCGCTACCTTTGGGCTGTGAGGTTC 113

Search completed: August 25, 2003, 23:47:55
Job time : 285.632 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 25, 2003, 19:49:37 ; Search time 2115.56 Seconds
(without alignments)
11683.734 Million cell updates/sec

Title: US-09-717-321A-15

Perfect score: 1017

Sequence: 1 cccctatctctgcagatt.....ccttggtctgtgagttc 1017

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

EST:
1: em_estbta:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estopl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_hmv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rtd:*
26: em_gss_phg:*
27: em_gss_vrt:*
28: gb_gss1:*
29: gb_gss2:*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	678.2	66.7	1796	11 AK081613	Mus muscu
2	677.6	66.6	2269	11 AK034601	Mus muscu
3	677.6	66.6	2270	11 AK047969	Mus muscu
4	660.2	64.9	2308	11 AK076023	Mus muscu

5	594.2	58.4	978	13 B0514423	B0514423 AGENCOURT
6	562	55.3	826	14 CA318577	B0318577 UI-M-FW0-
7	558.2	54.9	751	13 B0177713	B0177713 UI-M-ERO-
8	552.2	54.3	960	10 B0511766	B0511766 AGENCOURT
9	551.2	54.2	557	10 B665146	B665146 DRAB2A12
10	537.4	52.8	793	14 CA319273	CA319273 UI-M-FW0-
11	535.2	52.6	897	12 B1646446	B1646446 603279026
12	528.6	52.0	714	12 B1851489	B1851489 603377691
13	528.6	52.0	915	13 B0931374	B0931374 AGENCOURT
14	526	51.7	545	9 AA817948	AA817948 UI-R-AO-a
15	526	51.7	639	13 B0201343	B0201343 UI-R-D01-
16	525.4	51.7	760	13 B0191985	B0191985 UI-R-DN1-
17	510.4	50.2	806	13 B0444803	B0444803 UI-M-ERO-
18	508.2	50.0	515	14 CD372599	CD372599 UI-R-G00-
19	499.4	49.1	644	14 CA339407	CA339407 NISC_Ly02
20	493.8	48.6	686	14 CB057853	CB057853 NISC_Ly09
21	490.8	48.3	672	12 B1692509	B1692509 603343041
22	485	47.7	969	10 B783603	B783603 602109609
23	483.4	47.5	680	14 CD353147	CD353147 UI-M-G10-
24	482.8	47.5	741	12 B1694684	B1694684 603347620
25	482.4	47.4	732	14 CB596051	CB596051 AGENCOURT
26	479	47.1	502	9 A1028951	A1028951 UI-R-C0-1
27	474.8	46.7	787	12 B6968782	B6968782 602837287
28	473.4	46.5	705	13 B0200122	B0200122 UI-R-EB1-
29	472.4	46.5	687	12 BM389059	BM389059 UI-R-D20-
30	462.4	45.5	796	12 B1735385	B1735385 603356871
31	458.4	45.1	588	14 CA553516	CA553516 C0873C09-
32	454.2	44.7	805	12 B1692090	B1692090 603311752
33	453	44.5	584	12 CB055806	CB055806 NISC_J109
34	450.8	44.3	667	12 B0042534	B0042534 UI-M-EMO-
35	449.2	44.2	938	13 B0928530	B0928530 AGENCOURT
36	442.8	43.5	447	10 B663683	B663683 DRAB2A12
37	438.8	43.1	671	12 BM946306	BM946306 UI-M-EMO-
38	438.2	43.1	704	12 BM951935	BM951935 UI-M-ERO-
39	427.2	42.0	538	12 BM053597	BM053597 1468B05.Y
40	426.6	41.9	550	10 BM022481	BM022481 1461A04.Y
41	426.2	41.9	950	10 B168374	B168374 601774359
42	426	41.9	556	12 B1655954	B1655954 603285490
43	425.4	41.8	575	10 B1659623	B1659623 u288905.x
44	423.8	41.7	640	13 B0780117	B0780117 UI-R-FP0-
45	422	41.5	702	13 B0176893	B0176893 UI-M-D12-

ALIGNMENTS

RESULT 1
LOCUS AK081613
DEFINITION Mus musculus 16 days embryo head cDNA, RIKEN full-length enriched library, clone:Cl30051H24, product:RAS-related C3 botulinum substrate 1, full insert sequence.
ACCESSION AK081613
VERSION AK081613.1 GI:26349264
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y. High-efficiency full-length cDNA cloning by RT-PCR. *Nat. Biotechnol.* 17: 1273-1278 (1999)
2 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y. Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. *Genome Res.* 10 (10), 1617-1630 (2000)
JOURNAL MEDLINE
PUBMED
TITLE JOURNAL MEDLINE
PUBMED
11042159

REFERENCE
AUTHORS
3
Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
Konno, H., Akiyama, J., Nishi, K., Katsunai, T., Tashiro, H., Itoh, M.,
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,
Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
Fujisake, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Matsubara, M.,
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J.,
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A., and Hayashizaki, Y.
RIKEN Integrated sequence analysis (RISA) system -384-format
sequencing pipeline with 384 multicapillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)

JOURNAL
MEDLINE
PUBMED
11076861

TITLE
4
Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y.,
Aizawa, T., Hara, A., Fukunishi, Y., Konno, H., Adachi, J., Fukuda, S.,
Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamana, I.,
Saito, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Saito, R.,
Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T.,
Fleischmann, W., Gaasterland, T., Gissi, C., King, B., Kochiwa, H.,
Kuehl, P., Lewis, S., Matsuo, Y., Nikaido, I., Pesole, G.,
Quackenbush, J., Schriml, L.M., Staudli, F., Suzuki, R., Tomita, M.,
Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H.,
Baldarelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N.,
Carninci, P., de Bonaldo, M.F., Brownstein, M.J., Bult, C.,
Fletcher, C., Fujita, M., Cariboldi, M., Gustincich, S., Hill, D.,
Hofmann, M., Hume, D.A., Kamiya, M., Lee, N.H., Lyons, P.,
Marchionni, L., Mashima, J., Mazzarelli, J., Mombaerts, P., Nordone, P.,
Riing, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H.,
Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K.F., Suzuki, H.,
Toyo-Oka, K., Wang, K.H., Weitz, C., Whitaker, C., Wilting, L.,
Wynshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawai, H., Kohcenki, S.
and Hayashizaki, Y.
Functional annotation of a full-length mouse cDNA collection
Nature 409 (6821), 685-690 (2001)

JOURNAL
MEDLINE
PUBMED
11217851

TITLE
5
The FANTOM Consortium and the RIKEN Genome Exploration Research
Group Phase I & II Team.
Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)

JOURNAL
REFERENCE
AUTHORS
6 (bases 1 to 1796)
Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P.,
Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W.,
Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoaka, T., Hirozane, T.,
Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T.,
Kato, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M.,
Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,
Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N.,
Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N.,
Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,
Sogabe, T., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Kahira, S.,
Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,
Muramatsu, M., and Hayashizaki, Y.
Direct Submission
Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
Kanagawa 230-0045, Japan (E-mail: genome-res@gscc.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222,
Fax: 81-45-503-9216)
cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in RIKEN
Genomic Sciences Center and Genome Science Laboratory in RIKEN
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.
Please visit our web site for further details.
URL: http://genome.gsc.riken.go.jp/
URL: http://fantom.gsc.riken.go.jp/
Location/Qualifiers

FEATURES

source
1. 1796
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM:DB:C130051H24"
/db_xref="taxon:10090"
/clone="C130051H24"
/issue_type="head"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="16 days embryo"
1. 304
/note="unnamed protein product; PAS-related C3 botulinum
substrate 1 (MGI:MGI:97845, GB|AK01072, evidence: BLASTN,
100%, match=987)
putative"
/protein_id="BAC38272.1"
/db_xref="GI:26349265"
/translation="VRAKYPPEYRNHCPTPIIVTETKIDLRDDKDTIKLEKILTP
ITVQGLAKKEIGAVYDECSALTYGRKTYFDEAIRAVLCPPEPKRRKRCLL"
1781. 1786
/note="putative"
1796
polya_site
polya_signal
BASE COUNT 470 a 413 c 371 g 542 t
ORIGIN
Query Match 66.7%; Score 678.2; DB 11; Length 1796;
Best Local Similarity 85.3%; Pred. No. 1.7e-183;
Matches 871; Conservative 0; Mismatches 118; Indels 32; Gaps 9;
QY 1 CCCCTATCTGCTCGATTAAGAAATTCGCAAAATTCCTGTGTAAGTTCGCTG 60
DB 548 CCCCATTCTTCTGCTCGATTAAGAAATTCGCAAAATTCCTGTGTAAGTTCGCTG 607
QY 61 CTGAGAAACGTAAGACACTAAGCTGTGAGAGA---CTTGTGCTTAAGAAAGCTGAC 117
DB 608 CTGAGAAACGTAAGACACTAAGCTGTGAGAGACTTCTGTGTAAGAAAGCTGAC 667
QY 118 TTCTGGGCTCAGGGGTCAGACACCTCCCTGAGTCCAGACCGTGCAGACAGCAGCC 177
DB 668 TTCTGGGCTCAGGGGTCAGACACCTCCCTGAGTCCAGACCGTGCAGACAGCAGCC 727
QY 178 TCCTTAATGACAGCGTCCATGTAAAGCACTTAATCTATGACCCATGCTATTAAGT 237
DB 728 TCCTTAATGACAGCGTCCATGTAAAGCACTTAATCTATGACCCATGCTATTAAGT 237
QY 238 AACCTGTACTGACGTCAC---GATGGGCTGAACAGCTGCTGCTTGTGATTCATAGT 294
DB 788 AACCTGTACTGACGTCACGATGATGATGATGATGATGATGATGATGATGATGAT 847
QY 295 AGTTCTCTAAATACAGCTGACCGGCTTCGAGGCTTGAACAGAACTGCTGCTG 354
DB 848 AGTTCTCTAAATACAGCTGACCGGCTTCGAGGCTTGAACAGAACTGCTGCTG 354
QY 355 TGTTCCTCTAAGAAATATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 414
DB 908 TGTTCCTCTAAGAAATATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 414
QY 415 ACACGAGCTATCAAGAGAGACAGAGATTTTGAATTAAGAGTGAAGATTAATTT 474
DB 966 ACACGAGCTATCAAGAGAGAGAGAGATTTTGAATTAAGAGTGAAGATTAATTT 474
QY 475 ACACGAGCTATCAAGAGAGATTAATTAAGAGTGAAGAGTGAAGAGTGAAGAGT 534
DB 1026 ACACGAGCTATCAAGAGAGATTAATTAAGAGTGAAGAGTGAAGAGTGAAGAGT 1085
QY 535 ACAGTAAATTTCTGTAACAGTGAATGAATGCGGCTTATCATGCTGCTC----- 589
DB 1086 ACAGTAAATTTCTGTAACAGTGAATGAATGCGGCTTATCATGCTGCTGAG 1145
QY 590 -----TCTGAGCCCGCTTCCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 642
|| | |||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| |

Db 1146 CAYCCCTTGGCCCCCACCCTCCACATGCCGTCCTCCACAGATGGGCGTTGAGTCA 1205
 Oy 643 TATTAACTGGCCATCTCTACAGTTGCTTCTTGAAGTCTTTCTTGAAGACCCC 702
 Db 1206 TATTAACTGGCCATCTCTACAGTTGCTTCTTGAAGTCTTTCTTGAAGACCCC 1263
 Oy 703 TTTTAAACGACGATATGCTGACCTGATCTATTAAGATCTTTCTTGAAGTCTTT 756
 Db 1264 TTTTAAACGACGATATGCTGACCTGATCTATTAAGATCTTTCTTGAAGTCTTT 1323
 Oy 757 TCGGAGATTTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATG 816
 Db 1334 TTTGTT 1383
 Oy 817 -CTGACATGATGCTTCTCTGTTTCTGATGATGATGATGATGATGATGATGATG 875
 Db 1384 AGCTAGTATGATGCTTCTCTGTTTCTGATGATGATGATGATGATGATGATGATG 1443
 Oy 876 TTTACAGTTTTTAACTATTTTGAATATTTGAAACATGATGATGATGATGATGATG 935
 Db 1444 TTAGCGTTTTTAACTATTTTGAATATTTGAAACATGATGATGATGATGATGATG 1503
 Oy 936 ACTGTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 995
 Db 1504 ACTGTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1560
 Oy 996 T 996
 Db 1561 T 1561

RESULT 2
 Locus AK034601
 DEFINITION Mus musculus 12 days embryo embryonic body between diaphragm region and neck CDNA, RIKEN full-length enriched library, clone:9430013G17 product:RMS-related C3 botulinum substrate 1, full insert sequence.
 ACCESSION AK034601
 VERSION AK034601.1 GI:26330056
 KEYWORDS HTG; CAP trapper.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

REFERENCE
 AUTHORS Carninci, P. and Hayashizaki, Y.
 TITLE High-efficiency full-length cDNA cloning
 JOURNAL Meth. Enzymol. 303, 19-44 (1999)
 MEDLINE 99279253
 PUBMED 10349636
 REFERENCE
 AUTHORS Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
 TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
 JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
 MEDLINE 20493374
 PUBMED 11042159
 REFERENCE
 AUTHORS Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Kono, H., Akiyama, J., Nishi, K., Katsunari, T., Tashiro, H., Itoh, M., Suni, N., Ishii, Y., Nakamura, S., Hazama, M., Nishimura, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watanabe, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
 TITLE RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multiplexed sequencer
 JOURNAL Genome Res. 10 (11), 1757-1771 (2000)
 MEDLINE 20530913
 PUBMED 11076861
 REFERENCE
 AUTHORS Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y.,

Arakawa, T., Hara, A., Fukunishi, Y., Kono, H., Adachi, J., Fukuda, S., Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamanka, I., Saito, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Saito, R., Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T., Fleischmann, W., Gaasterland, T., Gissi, C., King, B., Kochwa, H., Kuehl, P., Lewis, S., Matsuo, Y., Nakai, T., Pesole, G., Quackenbush, J., Schriml, L. M., Staudt, F., Suzuki, R., Tomita, M., Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H., Balderelli, R., Barsh, D., Blake, J., Boffelli, D., Bojunga, N., Carninci, P., de Bonaldo, M. F., Brownstein, M. J., Bult, C., Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D., Hotmann, M., Hume, D. A., Kamiya, M., Lee, N. H., Lyons, P., Marchionni, L., Mashima, J., Mazzarelli, J., Mombauts, P., Nordone, P., Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H., Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K. F., Suzuki, H., Toyooka, K., Wang, K. H., Weitz, C., Whitaker, C., Wilmberg, L., Wynshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawai, J., Kohsaki, S. and Hayashizaki, Y.
 TITLE Functional annotation of a full-length mouse cDNA collection
 JOURNAL Nature 409 (6821), 685-690 (2001)
 MEDLINE 21085660
 PUBMED 11217851

REFERENCE
 AUTHORS The PANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
 TITLE Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
 JOURNAL Nature 420, 563-573 (2002)
 MEDLINE 1217851
 PUBMED 11217851

REFERENCE
 AUTHORS The PANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
 TITLE Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
 JOURNAL Nature 420, 563-573 (2002)
 MEDLINE 1217851
 PUBMED 11217851

TITLE
 JOURNAL
 COMMENT
 The PANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
 Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
 Nature 420, 563-573 (2002)
 URL: http://genome.gsc.riken.go.jp/

FEATURES
 source
 1..2269
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6J"
 /db_xref="PANTOM:DB:9430013G17"
 /db_xref="taxon:10090"
 /clone="9430013G17"
 /tissue_type="embryonic body between diaphragm region and neck"
 /clone_jib="RIKEN full-length enriched mouse cDNA library"
 /dev_stage="12 days embryo"
 198..776
 /note="unnamed protein product: RMS-related C3 botulinum substrate 1 [MGI:97845, GBAK01072, evidence: BLASTN, 100%, match=98]"

CDS

LOCUS AK076023 2308 bp mRNA linear HTC 07-DEC-2002
DEFINITION Mus musculus 10 days embryo whole body cDNA, RIKEN full-length
subtracted library, clone:261010016 product:RAS-related C3 botulinum
substrate 1, full insert sequence.
ACCESSION AK076023
VERSION AK076023.1 GI:26344957
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1
AUTHORS Carninci, P. and Hayashizaki, Y.
TITLE High efficiency full-length cDNA cloning
JOURNAL Meth. Enzymol. 303, 19-44 (1999)
MEDLINE 99279253
PubMed 10349636
REFERENCE 2
AUTHORS Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE 20499374
PubMed 11042159
REFERENCE 3
AUTHORS Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
Konno, H., Akiyama, J., Nishi, K., Kitsuami, T., Tashiro, H., Itoh, M.,
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A.,
Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Matsubara, M.,
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, D.,
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
TITLE RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer
JOURNAL Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE 20530913
PubMed 11076861
REFERENCE 4
AUTHORS Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y.,
Arai, T., Hara, A., Fukunishi, Y., Konno, H., Adachi, J., Fukuda, S.,
Aizawa, K., Izawa, M., Nishi, K., Kiyosaki, H., Kondo, S., Yamana, I.,
Saito, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Saito, R.,
Kadoya, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T.,
Fleischmann, W., Gaasterland, T., Gish, C., King, B., Kochiwa, H.,
Kuehl, P., Lewis, S., Matsuo, Y., Nikaido, I., Pesole, G.,
Wagner, L., Washio, T., Schiraldi, L. M., Staudt, F., Suzuki, R., Tomita, M.,
Magne, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H.,
Baldarelli, R., Barsh, G., Blake, J., Boffelli, D., Boljuga, N.,
Carninci, P., de Bonaldo, M. F., Brownstein, M. J., Bult, C.,
Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D.,
Hofmann, M., Hume, D. A., Kamiya, M., Lee, N. H., Lyons, P.,
Marchionni, L., Mashima, J., Mazzarelli, J., Mombauts, P., Nordone, P.,
Rang, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H.,
Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K. F., Suzuki, H.,
Toyo-Oka, K., Wang, K. H., Weitz, C., Whitaker, C., Wilming, L.,
Wynshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawai, H., Kohsaki, S.,
and Hayashizaki, Y.
TITLE Functional annotation of a full-length mouse cDNA collection
JOURNAL Nature 409 (6821), 685-690 (2001)
MEDLINE 21085660
PubMed 11217851
REFERENCE 5
AUTHORS The FANTOM Consortium and the RIKEN Genome Exploration Research
Group Phase I & II Team.
TITLE Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
JOURNAL Nature 420, 563-573 (2002)
MEDLINE 11217851
PubMed 11217851
REFERENCE 6 (phases 1 to 2308)
AUTHORS Adachi, J., Aizawa, K., Akahira, S., Akiyama, T., Aono, H., Arai, A.,
Arai, T., Bono, H., Carninci, P., Fukuda, S., Fukunishi, Y.,
Furuno, M., Hanagaki, T., Hara, A., Hayatsu, N., Ikegami, K.,
Ishikawa, T., Ishii, F., Imotani, K., Ishii, Y., Itoh, M., Izawa, M.,

kasukawa, T., Kato, H., Kawai, J., Kojima, Y., Konno, H., Kouda, M.,
Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Nishi, K.,
Nomura, K., Numazaki, R., Ohno, M., Okazaki, Y., Okido, T., Owa, C.,
Saito, H., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D.,
Shibata, K., Shibata, Y., Shinagawa, A., Shiraki, T., Sogabe, Y.,
Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F., Tanaka, T.,
Tejima, Y., Toya, I., Yamamura, T., Yamana, I., Yasunishi, A.,
Yoshida, K., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.
TITLE Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
Kanagawa 230-0045, Japan (E-mail: genome-res@gs.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222,
Fax: 81-45-503-9216)
COMMENT cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in Riken.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.
Please visit our web site for further details.
URL: http://genome.gsc.riken.go.jp/
URL: http://fantom.gsc.riken.go.jp/.
FEATURES
source
1..2308
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM,DB:261010016"
/db_xref="MGI:1898615"
/db_xref="taxon:10090"
/clone="261010016"
/tissue_type="whole body"
/clone_1kb="RIKEN full-length enriched mouse cDNA library"
/dev_stage="10 days embryo"
239..817
/note="unlabeled protein product; RAS-related C3 botulinum
substrate 1 (MSD) MGI:97845, GB|AK011072, evidence: BLASTN,
100%, match=987)
putative"
/codon_start=1
/protein_id="BAC36128.1"
/db_xref="GI:26344958"
/db_xref="MGI:97845"
/translation="MOAICVYVGDGAVKTCLLISYTNAPGEYIPYFDNYSNV
MVDGKPNVGLIMPTAGQEDYDRLPSPQTVFLICFSLVSAEENRAKYPYR
HRCPPNPIIVGKTLDRDQKDIETKIKLPIYVPGGLAAKELGAVKYLECSAL
TORGKLTVPDEAIRAVLCPPPVKKRRKCLLL"
2293..2298
/note="putative"
polyA_signal
2293..2298
polyA_site
2308
/note="putative"
BASE COUNT 559 a 580 c 518 g 650 t 1 others
ORIGIN
Query Match 64.9% Score 660.2; DB 11; Length 2308;
Best Local Similarity 84.8%; Pred. No. 3e-178;
Matches 866; Conservative 0; Mismatches 123; Indels 32; Gaps 10;
QY 1 CCCCTATTCCTTGGTCAGATTAGAAATTCGCAAAATCCTTCGTAAGTTCGCTTGC 60
DB 1061 CCCCATCTCTTGGTCAGATTAGAAATTCGCAAAATCCTTCGTAAGTTCGCTTGC 1120
QY 61 CTGAGAACAGTAAGCACTAAGCTGTGAGAGA---CTTGTCTCTTAAGAAAGCACTCAGC 117
DB 1121 CTGAGAACAGTAAGCACTAAGCTGTGAGAGACTTGTGCTGTAAGAAAGCACTCAGC 1180
QY 118 TTCTGG- GCTCAGGGGTGCAGACCTCCGCTCCAGACCCGTGTGACACGACACAGC 176
DB 1181 TTCTGGAAAGCAGGGGTGCAGACCTCCGCTCCAGACCCGTGTGACACGACACAGC 1240
QY 177 CTCCTTAATGACACGCTGCCATGTAAGCAGCACTGTAAGTATAGACCCATGCTCATTAAGC 236

```
Db      1241  CTCCTTAATGACACCTGTCATGTAACGACCTGTAATGTAACGCCATGTCATTTACG 1300
OY      237   TAACCTTGTACTGACGTCAC---GATGGGTGTAACACTGTGCTTGTGATTTCAATGT 293
Db      1301   TAACCTTGTACTGACGTCACGTAATGATGATGACACTGACCTGCTTGTGATTTCAATGT 1360
OY      294   GAGTTCCTTAAATACGAGTACGCGGCTTGTGACAGGCTTGAACAGAACTGTGACCT 353
Db      1361   GAGTTTCTTAAAGACGACGAGCTGTTTGTGCAACTTTGAACAGAACTGTGTTCT 1420
OY      354   GTTGTGCTCTGACGAGTATTTGTTCTTGTGCTGAGTGTGCTGCTGAGTGTGGA 413
Db      1421   GTGTGCTCTTAAAGAAATATTTGTTCTTGTGCTGAGTGTGCTGAGTGTGGA--GTGTGA 1478
OY      414   AACAGAGCTATCAAAAGAGACAGACGATTTTGTGATTAATGAAATAGAGTAAAT 473
Db      1479   AACAGAGATGATCAAAAGAGACAGACGATTTTGTGATTAATGAAATAGAGTAACT 1538
OY      474   TACACTACATTTGTACATGAGTAAATTCACATGAAATAAAGTGTACGGGTAAAGCTTTT 533
Db      1539   TACACTACATTTGTACATGAGTAAATTCACATGAAATAAAGTGTACGGGTAAAGCTTTT 1598
OY      534   AACGTTAATTTCTGTCAAAACAGTATGATGACAAATGCGCATCTTATGAGTGTCTC--- 589
Db      1599   AACGTTAATTTCTGTCAAAACAGTATGATGACAAATGCGCATCTTATGAGTGTCTGGA 1658
OY      590   -----TCTTGAAGCCCCCTTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 641
Db      1659   GCATCCCTTCCGCGCCCCCACCCTCCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1718
OY      642   ATATTTAAATCGCCATCCCTCAGTGTGCTTAACTTAACTTAACTTAACTTAACTTAACT 701
Db      1719   ATATTTAAATCGCCATCCCTCAGTGTGCTTAACTTAACTTAACTTAACTTAACTTAACT 1776
OY      702   CTCTTAAACGAGCAATATGCTGACGCTGCTTAACTTAACTTAACTTAACTTAACTTAACT 756
Db      1777   CTCTTAAACGAGCAATATGCTGACGCTGCTTAACTTAACTTAACTTAACTTAACTTAACT 1836
OY      757   TCGAGATATTTTGGTGTAGTATGTAAGAGTGGTTCCTTAACTTAACTTAACTTAACTTAACT 816
Db      1837   TTTTGTGTTTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1896
OY      817   -CTGACATGCTGCTTCCCTGCTTCTGTTTCTGTTTCTGTTTCTGTTTCTGTTTCTGTTTCT 875
Db      1897   AGCTAGTATGATGCTGCTTCTGTTTCTGTTTCTGTTTCTGTTTCTGTTTCTGTTTCTGTTT 1956
OY      876   TTACAGTTTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAACT 935
Db      1957   TTACAGTTTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAACT 2016
OY      936   ACTGTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 995
Db      2017   ACTGTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2073
OY      996   T 996
Db      2074   T 2074
```

```
RESULT 5
LOCUS   B0514423
DEFINITION AGNCODRT.10119699 NIH_MGC_134 Mus musculus cDNA clone
IMAGE:6511045 5', mRNA sequence.
ACCESSION B0514423
VERSION   B0514423.1 GI:22821949
KEYWORDS EST.
SOURCE   Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE 1 (bases 1 to 978)
```

```
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished
COMMENT  Contact: Robert Strausberg, Ph.D.
        Email: cgabs-remail.nih.gov
        Tissue Procurement: Dr. David Rowe
        cDNA Library Preparation: Invitrogen Corp
        DNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)
        DNA distribution by: Agencourt Bioscience Corporation
        Clone distribution: MGC clone distribution information can be
        found through the I.M.A.G.E. Consortium/LNLN at:
        http://image.llnl.gov
        Plate: LNL4080 row: 0 column: 14
        High quality sequence start: 17
        High quality sequence stop: 714.
        Location/Qualifiers
            1..978
                /organism="Mus musculus"
                /mol_type="mRNA"
                /db_xref="taxon:10090"
                /clone="IMAGE:6511045"
                /tissue_type="undifferentiated limb"
                /lab_host="DH10B (phage-resistant)"
                /note="Vector: pCMV-SPORT6.1.cdb; Site.1: Score; Site.2:
                NotI; Cloned unidirectionally. Primer: Oligo dT. Average
                insert size 1.7 kb. Constructed by Reagen, Invitrogen
                Corp. Note: this is a NIH_MGC Library."
BASE COUNT 249 a 224 c 201 g 301 t 3 others
ORIGIN
Query Match 58.4%; Score 594.2; DB 13; Length 978;
Best Local Similarity 85.8%; Pred. No. 1.9e-159;
Matches 762; Conservative 0; Mismatches 94; Indels 32; Gaps 8;
OY      1   CCCCTATTTCTGCTCAGATTAAGATTTGCCAAATACCTGTGATGATTAAGTTCGCTGTG 60
Db      70   CCCCATTTCTGCTCAGATTAAGATTTGCCAAATACCTGTGATGATTAAGTTCGCTGTG 129
OY      61   CTGANAACGTAACACTTAAGCTTTGAGAGA---CTTTGCTTAAAGAAAGCTGACGC 117
Db      130   CTGANAACGTAACACTTAAGCTTTGAGAGA---CTTTGCTTAAAGAAAGCTGACGC 189
OY      118   TTCTGGGCTCAGAGGGGTGAGACCCCTCCGCTGACCTCCAGACCGTGTGACAGCAGC 177
Db      190   TTCTGGACCCAGAGGGTGCAGACCCCTCCAGAGTTCCTCAGACCGTGTGACAGCAGC 249
OY      178   TCCCTTAATGACAGCGTGCATGTAAAGCAGCCTGTAACTTATCAGCCATGCTCAATACGT 237
Db      250   TCCCTTAATGACAGCGTGCATGTAAAGCAGCCTGTAACTTATCAGCCATGCTCAATACGT 309
OY      238   AACCTTGTACGTAGCTGAC---GATGGGTGTAACAGCTGTGCTTTGATTTCAATAGT 294
Db      310   AACCTTGTACGTAGCTGACGTGATGATGATGATGATGATGATGATGATGATGATGATGATG 369
OY      295   AGTTCCTAATAATACACAGTACCGGCTTGTGACAGCTTTTGAAGAGAACTGAGCTGCTG 354
Db      370   AGTTCCTAATAATACACAGTACCGGCTTGTGACAGCTTTTGAAGAGAACTGAGCTGCTG 429
OY      355   TGTGCTCTTAAAGAGAGTATTCCTGTTCTAGTGTGAGGCTGTGAGGAGTGTGAGAA 414
Db      430   TGTGCTCTTAAAGAGAGTATTCCTGTTCTAGTGTGAGGCTGTGAGGAGTGTGAGAA 487
OY      415   ACAGAGCTGATCAAAAGAGAGCAGATATTTGATCAATATGATGATGATGATGATGATGAT 474
Db      488   ACAGAGCTGATCAAAAGAGAGCAGATATTTGATCAATATGATGATGATGATGATGATGATG 547
OY      475   ACACATACATGATGAGTATGATCAATGATCAATGATCAATGATCAATGATCAATGATCAAT 534
Db      548   ACACATACATGATGAGTATGATCAATGATCAATGATCAATGATCAATGATCAATGATCAAT 607
OY      535   ACGGTTAATTTCTGTCAAAAGTATGATGATGATGATGATGATGATGATGATGATGATGATG 589
```


Db	ACGGTAAATTCCTGTCAAAACAGTAGGTGACAAATGGCCGATCTTATCATGTCTCTTGGAG	608
QY	-----TCCTGAGACCCCCCTTCCCCCGCTGTGCCCTCCACAGATGGGGGGCTTAGTCCA	590
Db	CATCCCTTGGCGCCGCCCCACCCCTCACCTGCGCCGTCCCCAGATGGGGGGCTTAGTCCA	668
QY	TATTTAAACTGGCCATCCTCAGTTGGTAACTTAGCAAGTCACTTCTTTTAGGACCCCC	643
Db	TATTTAAACCTGGCACCCCTCACAGTTCTTAACCTTAGCAAGTCTTTTCTTAG--ACCC	728
QY	TTCTTAAGACCAATATCTGTGACCTGTACATATAAGATCTTTCGTATATCA-----T	703
Db	TTCTTTACGACCAATATGTCTGTACCTGTACATATAAGATCTTTCGTATACCAANAGATT	786
QY	TCGGAGATTTTTTGTGTAGAT--AGTAGAACTGGCTGCTGTTTTCACACTCTCTTACTC	757
Db	TTTGTTTTTTTTTGGTAGTCAGTAGTAGAAGTGGCTCTGTTTTTGGCCCTACATTACTC	846
QY	--AACTGACTAGTCTTCCCTTCTGTTTTCTTAGTAACCTGGGGTGTAGAAA	815
Db	AGAGCTACTAGTGGCTTCTTAGTTTTCTTAACAACCTAGGTGGTGTACAA	906

RESULT 6
CALIBS77

LOCUS CAJ18577

DEFINITION CAl18577 826 bp mRNA linear EST 26-NOV-2002

ACCESSION U1-M-FW0-chr-o-14-0-U1_r1 NIH-BMAP_FW0 Mus musculus cDNA clone IMAGE:6813659 5', mRNA sequence.

VERSION CAl18577

KEYWORDS CAl18577.1 GI:24536701

SOURCE EST,

ORGANISM Mus musculus (house mouse)

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 826)
NIH-MGC http://mgc.ncl.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished

Contact: Robert Strausberg, Ph.D.
Email: gcrabbs@mail.nih.gov

Tissue Procurement: Dr. Jim Lin, University of Iowa
CDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:
http://image.lnl.gov

This clone was contributed by the Brain Molecular Anatomy Project (BMAP)

Seq primer: pyx-5.

```

FEATURES
source
Location/Qualifiers
1..826
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone_image="6813639"
/tissue_type="whole brain"
/dev_stage="embryo 13.5, 14.5, 16.5, 17.5dpc"
/lab_host="DH10B (T1 phage resistant)"
/clone_id="NH_BMAP_FW0"
/note="Organ: Brain; Vector: pYX-Asc; Site_1: EcoR I; Site_2: Not I; The library was constructed according Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. Denatured RNA was size fractionated on a 1% agarose gel. First strand cDNA synthesis was primed with oligo-dT primer containing a Not I site. Double strand cDNA was size selected according to mRNA size fraction, ligated with EcoR I adaptor, digested with NotI and then cloned directionally into pYX-Asc vector. The library tag sequence located between the Not I site and the polyA tail is AGCGAGACAG. This library was created for the University

```

BASE COUNT	226 a	195 c	162 g	242 t	1 others
ORIGIN	Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the Developing Mouse Nervous System', supported by National Institute of Mental Health (NIMH), Hemin Chin, Ph.D., program coordinator."				

Query Match	55.38;	Score 562;	DB 14;	Length 826;
Best Local Similarity	89.78;	Pred. No. 3.1e-150;		
Matches 677; Conservative	0;	Mismatches 56;	Indels 22;	Gaps 6

QY 1 CCCCTATTCGTGCAGATTAGAAATGGCAAAATCCTTGACACAAATGGGTGTG 60
 Db CCCCCTATTCGTGCAGATTAGAAATGGCAAAATCCTTGACACAAATGGGTGTG 135
 QY 61 CTGAGACACGTAAAGCCTTAAGCTGTGAGAGA--CTTGTGCTTAAGAGACTGCAGC 117
 Db CTTGAGACACGTAAAGCCTTAAGCTGTGAGAGACTCTTGTGCTTAAGAGACGGAGC 139
 QY 136 CTGAGAACACCTAAGCATTAACTCTCTTGAAGACTCTTGTGCTTAAGAGACGGAGC 139
 Db TTTGTGGGCTTAGGGGTGCAGACCCCTCCGTAAGCTCCAGACCGTGTGACACAGCAGCC 177
 QY 118 TTTGTGGGCTTAGGGGTGCAGACCCCTCCGTAAGCTCCAGACCGTGTGACACAGCAGCC 177

	178	179	180	181	182	183	184	185	186	187	188	189	190	191	192	193	194	195	196	197	198	199	200	201	202	203	204	205	206	207	208	209	210	211	212	213	214	215	216	217	218	219	220	221	222	223	224	225	226	227	228	229	230	231	232	233	234	235	236	237	238	239	240	241	242	243	244	245	246	247	248	249	250	251	252	253	254	255	256	257	258	259	260	261	262	263	264	265	266	267	268	269	270	271	272	273	274	275	276	277	278	279	280	281	282	283	284	285	286	287	288	289	290	291	292	293	294	295	296	297	298	299	300	301	302	303	304	305	306	307	308	309	310	311	312	313	314	315	316	317	318	319	320	321	322	323	324	325	326	327	328	329	330	331	332	333	334	335	336	337	338	339	340	341	342	343	344	345	346	347	348	349	350	351	352	353	354	355	356	357	358	359	360	361	362	363	364	365	366	367	368	369	370	371	372	373	374	375	376	377	378	379	380	381	382	383	384	385	386	387	388	389	390	391	392	393	394	395	396	397	398	399	400	401	402	403	404	405	406	407	408	409	410	411	412	413	414	415	416	417	418	419	420	421	422	423	424	425	426	427	428	429	430	431	432	433	434	435	436	437	438	439	440	441	442	443	444	445	446	447	448	449	450	451	452	453	454	455	456	457	458	459	460	461	462	463	464	465	466	467	468	469	470	471	472	473	474	475	476	477	478	479	480	481	482	483	484	485	486	487	488	489	490	491	492	493	494	495	496	497	498	499	500	501	502	503	504	505	506	507	508	509	510	511	512	513	514	515	516	517	518	519	520	521	522	523	524	525	526	527	528	529	530	531	532	533	534	535	536	537	538	539	540	541	542	543	544	545	546	547	548	549	550	551	552	553	554	555	556	557	558	559	560	561	562	563	564	565	566	567	568	569	570	571	572	573	574	575	576	577	578	579	580	581	582	583	584	585	586	587	588	589	590	591	592	593	594	595	596	597	598	599	600	601	602	603	604	605	606	607	608	609	610	611	612	613	614	615	616	617	618	619	620	621	622	623	624	625	626	627	628	629	630	
--	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	--

ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 751)
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Dr. James Lin, University of Iowa
CDNA Library preparation: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
This clone was contributed by the Brain Molecular Anatomy Project
(BMAP)
Seq primer: PYX-5.
Location/Qualifiers
1. 751
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone="IMAGE:5700194"
/tissue_type="whole brain"
/dev_stage="embryo 15.5 dpc"
/lab_host="DH10B (T1 phage resistant)"
/clone_lib="NIH-BMAP-ERO"
/note="Organ: Brain; Vector: pYX-asc; Site_1: EcoR I;
Site_2: Not I; The library was constructed according to
Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
1996. Denatured mRNA was size fractionated on a 1% agarose
gel. First strand cDNA synthesis was primed with an
oligo-dT primer containing a Not I site. Double stranded
cDNA was size selected according to mRNA size fraction,
ligated with EcoR I adaptor, digested with Not I, and then
cloned directionally into pYX-asc vector. The library tag
sequence located between the Not I site and the polyA tail
, is GTCGCGGAA. This library was created for the
University of Iowa Mouse Brain Molecular Anatomy Project
(BMAP). Gene Discovery in the Developing Mouse Nervous
System, supported by National Institutes of Mental Health
(NIMH), Hemin Chin, Ph.D., program coordinator."

BASE COUNT 198 a 180 c 154 g 217 t 2 others
ORIGIN

Query Match 54.98; Score 558.2; DB 13; Length 751;
Best Local Similarity 89.6%; Pred. No. 3.7e-149;
Matches 673; Conservative 0; Mismatches 55; Indels 23; Gaps 6;

QY 1 CCCCTATTCCTGTCGATTAAGAAATGCCAAATACCTTGTAACATAAGTTGGCTGTG 60
DB 6 CCCCCATTCCTGTCGATTAAGAAATGCCAAATACCTTGTAACATAAGTTGGCTGTG 65
QY 61 CTGGAACACGTAAGCACTAAGCTGTGAGAGA--CTTGTGCTTAAGAGATGACG 117
DB 66 CTGGAACACGTAAGCACTAAGCTGTGAGAGACTTCTGTGTAAGAGACCGCAGC 125
QY 118 TTCTGGGCTCAGGGGTCAGACCTCCCGTAGTCCAGACCGTGTGACAGACAGCC 177
DB 126 TTCTGGGCTCAGGGGTCAGACCTCCCGTAGTCCAGACCGTGTGACAGACAGCC 185
QY 178 TCCCTAATGACAGCTGCATGTAAGCACTGTAACCTATACCCCATGCTCATTTAGT 237
DB 186 TCCCTAATGACAGCTGCATGTAAGCACTGTAACCTATACCCCATGCTCATTTAGT 245
QY 238 AACTTTGACTGTAAGCTAC--GATGGGTGAACAGCTGCTTTGATTTGATTAAGT 294
DB 246 AACTTTGACTGTAAGCTACAGTATGATGTGACAGCTGCTTTGATTTGATTAAGT 305
QY 295 AGTTCTCTAAATACCAAGCTGACCGGCTTCTGACAGGCTTTGACAGAACTCTGGCTCTG 354

DB 306 AGTTCTCTAAATACCAAGCTGACCGGCTTCTGACAGGCTTTGACAGAACTCTGGCTCTG 365
QY 355 TGTTCCTCTAAGCAAGATATTCCTGCTAGCGCGGGGTGCGGGGTGAGTGTGAA 414
DB 366 TGTTCCTCTAAGCAAGATATTCCTGCTAGCGCGGGGTGCGGGGTGAGTGTGAA 423
QY 415 ACACGACCTCATCAAGAGAGACAGACAGTATTTGACATTAATATGAGATTAATTT 474
DB 424 ACACGACCTCATCAAGAGAGAGACAGACAGTATTTGACATTAATATGAGATTAATTT 483
QY 475 ACATGACATTTGACATGAGATTAATCAACGTGAATAAAGTGTGACGGTAAAGCTTTT 534
DB 484 ACATGACATTTGACATGAGATTAATCAACGTGAATAAAGTGTGACGGTAAAGCTTTT 543
QY 535 ACGGTTATTTCTGCAAAAGTATGATGACAAATGGCCGATCTTATGAGTGTTC 589
DB 544 ACGGTTATTTCTGCAAAAGTATGATGACAAATGGCCGATCTTATGAGTGTTC 603
QY 590 -----TCTTGAGCCCCCTTCCCTGCTGTCCTCCCGACAGATGGGCGTGAAGTCA 642
DB 604 CATCCCTTCGGGCCCCCGACCCCTCCACCTGCGCTCCTCCCGATGGGCGTGAAGTCA 663
QY 643 TATTTAAACGTGGCCATCTCTACAGTGTGCTAATAGCAAGTGTCTTTAGACCCCC 702
DB 664 TATTTAAACGTGGCCATCTCTACAGTGTGCTAATAGCAAGTGTCTTTAGAA--CCC 720
QY 703 TTCTTAAGCAAGCAATATGCTGACCTGACT 733
DB 721 TTCTTAAGCAAGCAATATGCTGACCTGACT 751

RESULT 8
BU511766 960 bp mRNA linear EST 12-SEP-2002
LOCUS AGENCOURT_10114251 NIH_MGC_134 Mus musculus cDNA clone
DEFINITION IMAGE:6506583 5', mRNA sequence.
BU511766
ACCESSION BU511766.1 GI:22817999
VERSION
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 960)
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Dr. David Rowe
CDNA Library Preparation: Invitrogen Corp
CDNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLAMA069 row: e column: 16
High quality sequence stop: 669.
Location/Qualifiers
1. 960
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:6506583"
/tissue_type="undifferentiated limb"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH-MGC_134"
/note="Vector: pCMV-SPORT6.1.cdb; Site_1: EcoR; Site_2:
NotI; Cloned unidirectionally. Primer: Oligo dr. Average
insert size 1.7 kb. Constructed by Resgen, Invitrogen
Corp. Note: this is a NIH-MGC Library."

BASE COUNT 246 a 218 c 196 g 298 t 2 others

ORIGIN				
Query Match	54.3%;	Score 552.2;	DB 13;	Length 960;
Best Local Similarity	86.6%;	Pred. No. 2.3e-147;		
Matches 733;	Conservative 0;	Mismatches 79;	Indels 34;	Gaps 10;
QY	45	ACCTAATGCTGCTGCTGAGAACGCTAAGCTAGCTGTTGAGAGAC---TTTGTC	101	
DB	3	AATTAATGCTGCTGCTGAGAACGCTAAGCTAAGCTGTTGAGAGACGCTGTTG	62	
QY	102	TTAAGAAGACTGAGCTTCTGGGCTCAGAGGCTGACAGCTCCGTAAGTCCAGACCGT	161	
DB	63	CTAAGAAGACGAGCTCTGAGAGCCAGAGGCTCAGACCTTTCAGAGTCCAGACCGT	122	
QY	162	GTGACACAGCAGCTCCTTAATGACAGCTGATGACAGCTGTAAGTAACTATACAG	221	
DB	123	GTGACACAGCAGCTCCTTAATGACAGCTGATGACAGCTGTAAGTAACTATACAG	182	
QY	222	CCCATGCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	278	
DB	183	CCCATGCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	242	
QY	279	CTTGAATTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	338	
DB	243	CTTGAATTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	302	
QY	339	AGAACTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	398	
DB	303	AGAACTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	362	
QY	399	GGGTGAGTGTGTAAGAACGAGCTGATGATGATGATGATGATGATGATGATGATGATG	458	
DB	363	GGGTGAGTGTGTAAGAACGAGCTGATGATGATGATGATGATGATGATGATGATGATG	420	
QY	459	AAGTGAAGATTAATTAACATGATGATGATGATGATGATGATGATGATGATGATGATG	518	
DB	421	AAGTGAAGATTAATTAACATGATGATGATGATGATGATGATGATGATGATGATGATG	480	
QY	519	CGGCTAAGAGTTTAAAGCTTAATTTCTCAAAAGATGATGATGATGATGATGATGATG	578	
DB	481	CGGCTAAGAGTTTAAAGCTTAATTTCTCAAAAGATGATGATGATGATGATGATGATG	540	
QY	579	ATCAGTGTCTC-----TCTGAGCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	636	
DB	541	ATCAGTGTCTCTGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	600	
QY	627	TGGGCGGTGAGTCAATTAATTAAGCTGAGCTGATGATGATGATGATGATGATGATGATG	686	
DB	601	TGGGCGGTGAGTCAATTAATTAAGCTGAGCTGATGATGATGATGATGATGATGATGATG	659	
QY	687	TTTCTTTAGAGCCCTCTCTTAAGCAGCAATATGCTGAGCTGATGATGATGATGATGATG	746	
DB	660	TTTCTTTAGAGCCCTCTCTTAAGCAGCAATATGCTGAGCTGATGATGATGATGATGATG	718	
QY	747	GATAATGCA-----TTCGAGATTTTGTGATAGTATGATGATGATGATGATGATGATG	798	
DB	719	GATAATGCA-----TTCGAGATTTTGTGATAGTATGATGATGATGATGATGATGATG	778	
QY	799	TCACCTCTCTTACTC--AGCTGAGTGTGCTTCCCTGCTT--TCTAGTAACTGGGCT	854	
DB	779	TCACCTCTCTTACTC--AGCTGAGTGTGCTTCCCTGCTT--TCTAGTAACTGGGCT	838	
QY	855	TAGAAA 860		
DB	839	TGCCAA 844		

RESULT 9
 LOCUS BG665146 557 bp mRNA linear EST 30-APR-2001
 DEFINITION DRABZ12 Rat DRG Library Rattus norvegicus cDNA clone DRABZ12 5',
 mRNA sequence.
 ACCESSION BG665146

VERSION				
KEYWORDS	EST.	BG665146.1	GI:13887068	
SOURCE	Rattus norvegicus (Norway rat)			
ORGANISM	Rattus norvegicus Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.			
REFERENCE	1 (bases 1 to 557) Xiao, H.S., Huang, Q.H., Zhang, F.X., Bao, L., Lu, Y.J., Guo, C., Yang, L., Huang, W.J., Fu, G., Xu, S.H., Cheng, X.P., Yan, Q., Zhu, Z.D., Zhang, X., Chen, Z., Han, Z.G., and Zhang, X. Identification of gene expression profile of dorsal root ganglion in the rat peripheral axotomy model of neuropathic pain Proc. Natl. Acad. Sci. U.S.A. 99 (12), 8360-8366 (2002)			
TITLE	JOURNAL MEDLINE PUBMED			
COMMENT	12060780 Contact: Zhang Xu Laboratory of Sensory System Institute of Neuroscience 320 Yue Yang Road, Shanghai 200031, P. R. China Tel: 86-21-64748700-121 Fax: 86-21-64713446 Email: xu.zhang@ion.ac.cn This clone is also available at Chinese National Human Genome Center at Shanghai, 351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong New Area, P.R.China. Please contact with Zhang Xu (xu.zhang@ion.ac.cn) or Han Zeguang (hanzg@chgc.sh.cn) PCR Primers FORWARD: T3 BACKWARD: T7 Seq primer: T3 POLYA-No.			
FEATURES				
source				
1..557 /organism="Rattus norvegicus" /mol_type="mRNA" /strain="Sprague-Dawley" /db_xref="taxon:10116" /clone="DRABZ12" /sex="male" /tissue_type="dorsal root ganglion" /dev_stage="adult" /clone_lib="Rat DRG Library"				
BASE COUNT 142 a 130 c 128 g 156 t 1 others				
ORIGIN				
Query Match 54.2%; Score 551.2; DB 10; Length 557; Best Local Similarity 99.3%; Pred. No. 3.2e-147; Matches 553; Conservative 0; Mismatches 4; Indels 0; Gaps 0;				
QY	55	GTTGTGCTGAGAACGCTAAGCACTAAGCTGTTGAGAGACTTGTCTTAAGAAAGCTGC	114	
DB	1	GTTGTGCTGAGAACGCTAAGCACTAAGCTGTTGAGAGACTTGTCTTAAGAAAGCTGC	60	
QY	115	AGCTTCTGGGCTCAGAGGCTCAGACCTCCCGTAAGTCCCAAGCCGTGTACACAGCACA	174	
DB	61	AGCTTCTGGGCTCAGAGGCTCAGACCTCCCGTAAGTCCCAAGCCGTGTACACAGCACA	120	
QY	175	GCCTCCTTAATGACAGCTGCATGTAAGCAGCTGTAATTAATTAATTAATTAATTAATTA	234	
DB	121	GCCTCCTTAATGACAGCTGCATGTAAGCAGCTGTAATTAATTAATTAATTAATTAATTA	180	
QY	235	CGTAACTTTGTAAGTCAAGTCAAGTGGGTGAACAGCTGCTCTTGAATTCATGATG	294	
DB	181	CGTAACTTTGTAAGTCAAGTCAAGTGGGTGAACAGCTGCTCTTGAATTCATGATG	240	
QY	295	AGTTCTCTAAATATACAGCTGACCGGCTTGTGAGGCTTTGAACAGAACTCTGGCTCG	354	
DB	241	AGTTCTCTAAATATACAGCTGACCGGCTTGTGAGGCTTTGAACAGAACTCTGGCTCG	300	
QY	355	TGTTGCTCTAAGCAAGATTTCTGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTG	414	
DB	301	TGTTGCTCTAAGCAAGATTTCTGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTG	360	

QY 415 ACAGACGTCATCAAGAGAGACAGACAGTATTTGACTAATATGAGATTAATTT 474
DB 361 ACAGAGCTCATCAAGAGAGACAGACAGTATTTGACTAATATGAGATTAATTT 420
QY 475 ACACACTATTTACATGAGATTAATTCAGATTAATAAGTTCACGGGTAAAGCTTTTAA 534
DB 421 ACACACTATTTACATGAGATTAATTCAGATTAATAAGTTCACGGGTAAAGCTTTTAA 480
QY 535 ACAGGTATTTCTGCAACAGATGATGACAAATGCGCATCTATCATGCTCTCTCTG 594
DB 481 ACAGGTATTTCTGCAACAGATGATGACAAATGCGCATCTATCATGCTCTCTCTG 540
QY 595 AGCCCCCTTCCCTG 611
DB 541 AGCCCCCTTCCCTG 557

RESULT 10
CA319273 793 bp mRNA linear EST 26-NOV-2002
LOCUS UI-M-FM0-cbz-a-24-0-UI.r1 NIH_BMAP_FM0 Mus musculus cDNA clone
DEFINITION IMAGE: 6816385 5', mRNA sequence.
ACCESSION CA319273.1 GI:24537397
VERSION CA319273.1 GI:24537397
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
TITLE NIH-MGC http://mgc.ncl.nih.gov/.
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgabbs@mail.nih.gov
Tissue Procurement: Dr. Jim Lin, University of Iowa
CDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
This clone was contributed by the Brain Molecular Anatomy Project
(BMAP)

Seq primer: PYX-5.
This clone was contributed by the Brain Molecular Anatomy Project
(BMAP)

FEATURES
source location/Qualifiers

1..793
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone="IMAGE: 6816385"
/tissue_type="whole brain"
/dev_stage="embryo 13.5,14.5,16.5,17.5dpc"
/lab_host="DH10B (T1 phage resistant)"
/clone_11b="NIH_BMAP_FM0"
/note="Organ: Brain; Vector: PYX-Asc; Site:1: EcoR I;
Site_2: Not I; The library was constructed according
Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
1996. Denatured RNA was size fractionated on a 1% agarose
gel. First strand cDNA synthesis was primed with oligo-dT
primer containing a Not I site. Double strand cDNA was
size selected according to mRNA size fraction, ligated
with EcoR I adaptor, digested with NotI and then cloned
directionally into PYX-Asc vector. The library tag
sequence located between the Not I site and the polyA tail
is AGCAGACAG. This library was created for the University
Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the
Developing Mouse Nervous System', supported by National
Institute of Mental Health (NIMH), Hemlin Chin, Ph.D.,
Program coordinator."

200 a 185 c 165 g 242 t 1 others

ORIGIN

Query Match 52.8%; Score 537.4; DB 14; Length 793;
Best Local Similarity 87.2%; Pred.No. 3.7e-143;
Matches 677; Conservative 0; Mismatches 72; Indels 27; Gaps 7;

QY 64 AGAACACCTAAGCAGTAACTGTTGAGAG--CTTGTCTCTTAAGAGAGTGCAGCTTC 120
DB 1 AGAACACCTAAGCAGTAACTGTTGAGAGAGCTTCTTGTGTAAGAGACCGCAGCTTC 60
QY 121 TGGGCTCAGGGGTGACAGACCTCCGTAAGCTCCAGACCTGTGTGACAGACAGACCTCC 180
DB 61 TGGAGCCAGGGGTGACAGACCTCCGTAAGCTCCAGACCTGTGTGACAGACAGACCTCC 120
QY 181 TTAATGACAGCGTGCATGTAACGACCTGTAACCTATACAGCCCATGTCATTAAGCTAC 240
DB 121 TTAATGACAGCGTGCATGTAACGACCTGTAACCTATACAGCCCATGTCATTAAGCTAC 180
QY 241 TTTGTACTGTACGTCAC--GATGGGTAAACAGCTGCTCTTGTGATTTGATAGTAGT 297
DB 181 TTTGTACTGTACGTCACGATGATGAGTGTGACAGCTGCTCTTGTGATTTGATAGTAGT 240
QY 298 TCTCTAAATACACGCTACCGGCTTGTGACAGGCTTTGACAGAACTCTGCTCTGTGT 357
DB 241 TTTCTAAATACACGCTACCGGCTTGTGACAGGCTTTGACAGAACTCTGCTCTGTGT 300
QY 358 TGGCTCTAAGCAGATATCTGTCCTAGTGGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 417
DB 301 TGGCTCTAAGCAGATATCTGTCCTAGTGGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 358
QY 418 CGAGCTCATCAAGAGAGAGACAGATATTTGACAAATATGAGATTAATTTACA 477
DB 359 CGAGCTCATCAAGAGAGAGACAGATATTTGACAAATATGAGATTAATTTACA 418
QY 478 CTACATTTACATGAGATTAATTCAGCTGAATTAAGTTCACAGGCTAAGCTTTTAAAG 537
DB 419 CTACATTTACATGAGATTAATTCAGCTGAATTAAGTTCACAGGCTAAGCTTTTAAAG 478
QY 538 GTTATTTCTGCAACAGTATGATGACAAATGGCGGATCTATCGTGTCTC----- 589
DB 479 GTTATTTCTGCAACAGTATGATGACAAATGGCGGATCTATCGTGTCTC----- 538
QY 590 ---TCTTGAGCCCTTCCCTCCCTGCTGCTCCCTCCAGATGGGGCTGAGTCCATAT 645
DB 539 CCGCTCGGCGCCCGCCACCTCCACCTCCAGTGGCGGCTGAGTCCATAT 598
QY 646 TTAACCTGGCATCTCTCAGCTGCTAATTAAGAGTCTTTTATAGACCCCTTC 705
DB 599 TTAACCTGGCATCTCTCAGCTGCTAATTAAGAGTCTTTTATAGACCCCTTC 656
QY 706 TTAACGACCAATATGCTGACCTGACTATTAAGATCTTCTGATTAATGCA-----TTCGG 760
DB 657 TTAACGACCAATATGCTGACCTGACTATTAAGATCTTCTGATTAATGCA-----TTCGG 716
QY 761 AGATTTTGTGATAGATGTAAGTGGTCTGTTTCTTACCTCTCTTACTGAG 816
DB 717 GTTTTGTGATAGATGTAAGTGGTCTGTTTCTTACCTCTCTTACTGAG 772

RESULT 11
B1646446 897 bp mRNA linear EST 12-SEP-2001
LOCUS B1646446 6032790261 NCI_CGAP_Mam3 Mus musculus cDNA clone IMAGE:5319127 5',
DEFINITION mRNA sequence.
ACCESSION B1646446.1 GI:15560682
VERSION B1646446.1 GI:15560682
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NIH-MGC http://mgc.ncl.nih.gov/.

TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL	Unpublished
COMMENT	Contact: Robert Strausberg, Ph.D.

FEATURES	Location/Qualifiers
source	1. .897

BASE COUNT	228 a	213 c	190 g	266 t
ORIGIN				

Oy	590	-----TCTTGAGCCCCCCTTCCTCCCGCTGTCCTCCACAGATGGGGCGTTGAATCC	64.1
Db	538	GCAATCCCTTCGGGCCCCCACCCTCACCTGCAGTCGCCTCCGCCAATAGGGGGCGATGATCC	59.7
Oy	642	ATAATTAAACT--GGCATCTCACAAGTTGCCTAACCTAGACAGTCCTTTCTTAGAACCC	70.0
Db	598	ATATTTAACCTGGGGCCACCCCTCACAGTGTCTTAACCTAGCAAGTCTTTCTTTTA--GAAC	65.6
Oy	701	CCATTCTTAACGACCAATATGTCTGACCTGACCTAATAATCTTCTGATATGA----T	75.6
Db	657	CCTTCTTAAACGACCAATATGTCTGACCTGACCTAATAATCTTCTGATATGAAGA	71.6
Oy	757	TCCGAGATTTTTTTGGTAGATAGTAGAAGTGGTTCCTGTTTACCTCTCCTTTACTACAG	81.6
Db	717	TTTGGTTCCTCCTTGGTAGCTCAGTAAGAAATGGTGCTCTTTTTCGCTTACTTACTACAG	77.6
Oy	817	-CTGACTAGTCTCCCTTCGTTTTCTAGTAACATGGGTGT--AGAAATCAAGCTGTACG	87.2
Db	777	AGCTAGTAAAGTGTCCCTTAGTTTCTCCAGCAAATAGAGTGTGGCAATATGTGTGGC	83.6
Oy	873	GCTTTACAGTTTTTAACTAATTATTAGATATCTGAAC	91.0
Db	837	CGCCTTACGGGTTTAACTAATTAGATATCTTAAAC	87.4

RESULT 12	
B1851489	
LOCUS	B1851489 714 bp mRNA linear EST 10-Oct-2000
DEFINITION	603377651f1 NCI_CGAP_Man2 Mus musculus cDNA clone IMAGE:5390207 5'
	mRNA sequence.
ACCESSION	B1851489
VERSION	B1851489.1 GI:15992236
KEYWORDS	EST.
SOURCE	Mus musculus (house mouse)
ORGANISM	Mus musculus
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 714) NTH-MGC http://mgc.ncl.nih.gov/ National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished Contact: Robert Strausberg, Ph.D.
REFERENCE	
TITLE	
JOURNAL	
COMMENT	

FEATURES	Location/Qualifiers
source	1. .714

db 478 AACGTTAATTCTCTCAACAGTAGGTGACAAATGCCCGAATCTTATCAGTGTCTTGA 537

Oy	814	CAGCTG--ACNAGTGCCTCCCTTCGTTTCTAGTACCGGGGTGAGAAATCAGCTGCG	874
Db	726	CAGAACGCTAGTACTGCTCTCTCTAGTTCTTAGGCACTACTAGCTGCACAATCATGTGTTCC	785
Oy	872	GGCTTTACA---GTTTTTAAACTATTTTATGATTCGAAA	909
Db	786	CAGCTTTTACGGGTTTTTAAACTATTTTATGAAATTCCTTAA	826
RESULT 14			
AA817948/c			
LOCUS			
DEFINITION			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PubMed			
COMMENT			
On Feb 17, 1998 this sequence version replaced gl:2887828.			
Contact: Soares, MB			
Coordinated Laboratory for Computational Genomics			
University of Iowa			
375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA			
Tel: 319 335 8250			
Fax: 319 335 9565			
Email: bento-soares@uiowa.edu			
The sequence tag present in the cDNA between the NotI site and the			
oligo-dT track served to identify it as a clone from the normalized			
adult kidney library. cDNA Library Preparation: M. Palma Bonaldo,			
Ph.D. Clone distribution: clones will be available through Research			
Genetics This clone is also available through the I.M.A.G.E.			
Consortium at LNL (info@image.lnl.gov). IMAGE ID=167351			
Seq primer: M13 Forward			
POLYA-No.			
FEATURES			
source			
1..545			
Location/Qualifiers			
/organism="Rattus norvegicus"			
/mol_type="mRNA"			
/strain="Sprague-Dawley"			
/db_xref="taxon:10116"			
/clone="UI-R-A0-ag-e-05-0-UI"			
/dev_stage="adult"			
/lab_host="DH10B (Life Technologies)"			
/note="Vector: pRT3D-Pac (Pharmacia) with a modified			
polylinker. Site.1: Not I; Site.2: Eco RI; This library			
consists of a mixture of individually tagged normalized			
libraries constructed from rat placenta, adult lung, brain			
, liver, kidney, heart, spleen, ovary, and muscle. The tag			
is a string of 3-5 nucleotides present between the Not I			
site and the oligo-dT track which allows identification of			
the library of origin of a clone within the mixture."			
BASE COUNT			
ORIGIN			
148 a 120 c 119 g 158 t			
Query Match			
Best Local Similarity 100.0%; Pred. No. 57e-140;			
Matches 526; Conservative 0; Mismatches 0; Indels 0; Gaps 0;			
1 CCCCTATTCTTGCTCAGTAAAGATTGCCAAAATACCTGTGACTAAGTTGGCTGTG 60			

DB	SEQ ID	DEFINITION	LOCUS	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	MEDLINE	PUBMED	COMMENT
Db	540	CCCGTATCTTGTGCAGATTGAAGATTGGCCAAATTAACCTTGTGAACCTAAGTTGGTGTG													
QY	61	CTGAGAACACGTTAAGCACTAAGCTGTGGAGAGACTTTGTCTTTAAGAAGACTGCAGCTTC	BQ201343	BQ201343	UI-R-DQ1-clt-n-08-0-UI.s1	UI-R-DQ1 Rattus norvegicus cDNA clone									
Db	480	CTGGAACACGTTAAGCACTAAGCTGTGGAGAGACTTTGTCTTTAAGAAGACTGCAGCTTC		BQ201343	UI-R-DQ1-clt-n-08-0-UI 3', mRNA sequence.										
QY	121	TGGGCTCAGGGGTGCAGACCTCCCGTAGTCCACAGACCGTGTGACACAGCAGACCTCC		BQ201343.1	GI:20417808	EST.									
Db	420	TGGGCTCAGGGGTGCAGACCTCCCGTAGTCCACAGACCGTGTGACACAGCAGACCTCC													
QY	181	TTAATGACAGCTGCCATGTAAGCACCCTGTAACCTATACAGCCCATGCTCATTAACGTAAC													
Db	360	TTAATGACAGCTGCCATGTAAGCACCCTGTAACCTATACAGCCCATGCTCATTAACGTAAC													
QY	241	TTTGACTGTCAGTCAGCATGGGTGTGAACACCTCGCTTGAATTCATGTGAGATTC													
Db	300	TTTGACTGTCAGTCAGCATGGGTGTGAACACCTCGCTTGAATTCATGTGAGATTC													
QY	301	CTAAATACACAGCTGACCGGCTTGTGCAGGCTTTGAAACAGAACTCTGGCTCTGTGTGC													
Db	240	CTAAATACACAGCTGACCGGCTTGTGCAGGCTTTGAAACAGAACTCTGGCTCTGTGTGC													
QY	361	CTCTACAGAGATTTCTTGTCTTCTAGTGTGGGTGCTGGGTGGAGTGTGAAACACGA													
Db	180	CTCTACAGAGATTTCTTGTCTTCTAGTGTGGGTGCTGGGTGGAGTGTGAAACACGA													
QY	421	CGTATCAAGAGACAGACAGATATTTTGCATTAATGAAAGTATGAGATTAATTATACACTA													
Db	120	CGTATCAAGAGACAGACAGATATTTTGCATTAATGAAAGTATGAGATTAATTATACACTA													
QY	481	CATTGTACATGGAGTAATTCACATGAAATAAAGTGTCAAGGGTAA													
Db	60	CATTGTACATGGAGTAATTCACATGAAATAAAGTGTCAAGGGTAA													

RESULT 15
BQ201343/c

BQ201343 639 bp mRNA linear EST 02-MAY-2002

UI-R-DQ1-clt-n-08-0-UI.s1 UI-R-DQ1 Rattus norvegicus cDNA clone

UI-R-DQ1-clt-n-08-0-UI 3', mRNA sequence.

BQ201343

BQ201343.1 GI:20417808

EST.

Rattus norvegicus (Norway rat)

Rattus norvegicus

Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

1 (bases 1 to 639)

Bonaldi M.F., Lennon G. and Soares M.B.

Normalization and subtraction: two approaches to facilitate gene discovery

Genome Res. 6 (9), 791-806 (1996)

9704447

8889548

Contact: Soares, MB

Coordinated Laboratory for Computational Genomics

University of Iowa

375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA

Tel: 319 335 8250

Fax: 319 335 9565

Email: bent-soares@uiowa.edu

The sequence contained an oligo-dT track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. The sequence tag present in the cDNA between the NotI site and the oligo-dT track served to verify it as a clone from the normalized rat cell line R3327-5a library cDNA library preparation::

M.B. Soares Lab Clone distribution: clones will be available through Research Genetics (www.resgen.com)

Seq primer: M13 Forward

POLYA=yes.

FEATURES
sourceLocation/Qualifiers
1. 639

/organism="Rattus norvegicus"

/mol_type="mRNA"

/strain="Sprague-Dawley"

/db_xref="taxon:10116"

/clone="UI-R-DQ1-clt-n-08-0-UI"

/dev_stage="adult"

/lab_host="DH10B (Life Technologies)"

/clone_lib="UI-R-DQ1"

/note="vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-DQ1 library is a normalized Rat Cell Line R3327-5A library (nr5a) constructed in pT37 PAC vector according to the procedure described by Bonaldo, Lennon & Soares (Normalization and Subtraction: Two Approaches to Facilitate Gene Discovery. Genome Research 6: 791-806, 1996). The oligonucleotide used to prime first strand synthesis contained the sequence tag GGACTAGATC between the Not I cloning site and dt18 stretch. The Rat Cell Line R3327-5A was provided by Mary Hendrix of the University of Iowa.

TAG_LIB="UI-R-DQ1"
TAG_TISSUE="rat cell line R3327-5a"
TAG_SEQ="GGACTAGATC"BASE COUNT 183 a 130 c 136 g 190 t
ORIGIN

Query Match

51.7%; Score 526; DB 13; Length 639;

Best Local Similarity 100.0%; Pred. NO. 6.3e-140;

Matches 526; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 CCCCTATCTTGCCTCAGATTAGAAATGCGAAATACCTGTGAACTAAGTGGCTGTG 60
DB 541 CCCCTATCTTGCCTCAGATTAGAAATGCGAAATACCTGTGAACTAAGTGGCTGTG 482
QY 61 CTGAGACACGTAGCCTAAGCTGTAGAGAGCTTTCCTTAAGAAGACTGCAGCTTC 120
DB 481 CTGAGACACGTAGCCTAAGCTGTAGAGAGCTTTCCTTAAGAAGACTGCAGCTTC 422
QY 121 TGGGCTCAGGGGTGAGACCCCTCCGCTAGCTCCAGACCGCTGTGACACAGCAGCTCC 180
DB 421 TGGGCTCAGGGGTGAGACCCCTCCGCTAGCTCCAGACCGCTGTGACACAGCAGCTCC 362
QY 181 TTAATGACACGCTGCCATGTATACGACCTGTAACTATACAGCCATGCTATTACGTAC 240
DB 361 TTAATGACACGCTGCCATGTATACGACCTGTAACTATACAGCCATGCTATTACGTAC 302
QY 241 TTTGTACTGTACGTACGATGGGGTTAAACAGCTCTGCTTTGATTTCATAGTGAATCT 300
DB 301 TTTGTACTGTACGTACGATGGGGTTAAACAGCTCTGCTTTGATTTCATAGTGAATCT 242
QY 301 CTAAATATACAGCTACCGGCTTTCGACGGCTTTAAGACAACTCTGGCTCTGTGTTC 360
DB 241 CTAAATATACAGCTACCGGCTTTCGACGGCTTTAAGACAACTCTGGCTCTGTGTTC 182
QY 361 CTCTAACGAAGATCTCTGCTTCTAGTCGAGGGTGTGCGGGTGTGAGAGTGTGAACACGA 420
DB 181 CTCTAACGAAGATCTCTGCTTCTAGTCGAGGGTGTGCGGGTGTGAGAGTGTGAACACGA 122
QY 421 CGTCATVCAAGAGACAGACAGTATTTGACTAATATGAAAGTAGAGATTAAATTTACACTA 480
DB 121 CGTCATVCAAGAGAGACAGTATTTGACTAATATGAAAGTAGAGATTAAATTTACACTA 62
QY 481 CATTGTACATGAGATTAATTCAGTAAAGTGTCAAGGGTAAA 526
DB 61 CATTGTACATGAGATTAATTCAGTAAAGTGTCAAGGGTAAA 16
```

Search completed: August 25, 2003, 23:35:36
Job time : 2120.56 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using SW model

Run on: August 25, 2003, 21:58:43 ; Search time 74.6032 Seconds
(without alignments)
6016.990 Million cell updates/sec

Title: US-09-717-321a-15

Perfect score: 1017
Sequence: 1 cccctattctgtcagatt.....ccttggctgtgagttc 1017

Scoring table: IDENTITY_NTC
Gapop 10.0 , Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-Processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

1: Issued_Patents_NA:*
2: /cgn2_6/prodata/1/ina/5A_COMB.seq:*
3: /cgn2_6/prodata/1/ina/5B_COMB.seq:*
4: /cgn2_6/prodata/1/ina/5A_COMB.seq:*
5: /cgn2_6/prodata/1/ina/5B_COMB.seq:*
6: /cgn2_6/prodata/1/ina/5A_COMB.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	187.2	18.4	605	3	US-09-385-982-483	Sequence 483, App
C 2	68	6.7	588	3	US-09-385-982-133	Sequence 133, App
C 3	54.2	5.3	7218	1	US-08-232-463-14	Sequence 14, App1
C 4	35.4	3.5	2038	1	US-08-181-271A-1	Sequence 1, App1
C 5	35.4	3.5	2038	1	US-08-449-315-1	Sequence 1, App1
C 6	35.4	3.5	2038	1	US-08-449-803-1	Sequence 1, App1
C 7	35.4	3.5	2038	1	US-08-449-043-1	Sequence 1, App1
C 8	35.4	3.5	2038	1	US-08-456-265A-1	Sequence 1, App1
C 9	35.4	3.5	2038	1	US-08-455-416-1	Sequence 1, App1
C 10	35.4	3.5	2038	1	US-08-455-244-1	Sequence 1, App1
C 11	35.4	3.5	2038	1	US-08-454-876-1	Sequence 1, App1
C 12	35.4	3.5	2038	2	US-08-457-364-1	Sequence 1, App1
C 13	35.4	3.5	2038	2	US-08-456-262-1	Sequence 1, App1
C 14	35.4	3.5	2038	2	US-08-456-240-1	Sequence 1, App1
C 15	35.4	3.5	2038	2	US-08-455-736-1	Sequence 1, App1
C 16	35.4	3.5	2038	2	US-08-971-211-1	Sequence 1, App1
C 17	35.4	3.5	2038	2	US-09-350-600-1	Sequence 1, App1
C 18	34	3.3	5084	1	US-08-306-691B-25	Sequence 21, App1
C 19	34	3.3	5084	5	PCT-US93-06251-21	Sequence 25, App1
C 20	33.2	3.3	15393	4	US-09-453-702B-181	Sequence 191, App
C 21	32.2	3.2	453	4	US-09-484-970B-85	Sequence 85, App
C 22	32	3.1	43360	4	US-09-453-702B-206	Sequence 206, App
C 23	32	3.1	45325	4	US-09-453-702B-261	Sequence 261, App
C 24	31.6	3.1	2703	4	US-09-482-273-75	Sequence 75, App1
C 25	31.6	3.1	2709	4	US-09-482-273-75	Sequence 101, App
C 26	31.6	3.1	3001	4	US-09-539-333D-158	Sequence 158, App
C 27	31.6	3.1	3001	4	US-09-539-333D-160	Sequence 160, App

28	31.6	3.1	4203	2	US-08-866-757-1	Sequence 1, App1
29	31.6	3.1	4203	3	US-09-153-593-1	Sequence 1, App1
C 30	31.2	3.1	2559	4	US-09-569-098A-109	Sequence 109, App
C 31	31	3.0	1221	4	US-09-134-001C-365	Sequence 365, App
C 32	30.8	3.0	946	4	US-09-599-360B-42	Sequence 42, App
C 33	30.6	3.0	684	4	US-09-134-001C-1999	Sequence 1999, App
C 34	30.6	3.0	952	4	US-09-174-768-2	Sequence 2, App1
C 35	30.6	3.0	58407	4	US-08-916-421B-2	Sequence 2, App1
C 36	30.4	3.0	622	3	US-09-109-204-9	Sequence 9, App1
C 37	30.4	3.0	632	4	US-09-490-032-9	Sequence 9, App1
C 38	30.4	3.0	1422	4	US-09-227-957-18	Sequence 18, App1
C 39	30.4	3.0	1975	3	US-09-109-204-4	Sequence 4, App1
C 40	30.4	3.0	1975	4	US-09-490-032-4	Sequence 4, App1
C 41	30.4	3.0	3010	4	US-09-996-243-313	Sequence 313, App
C 42	30.4	3.0	87350	4	US-08-781-891-79	Sequence 79, App1
C 43	30.4	3.0	87350	4	US-09-618-166-79	Sequence 79, App1
C 44	30.4	3.0	87343	4	US-09-791-211-3	Sequence 3, App1
C 45	30.2	3.0	13865	3	US-09-009-217-11	Sequence 11, App1

ALIGNMENTS

```
RESULT 1
US-09-385-982-483/C
; Sequence 483, Application US/09385982
; Patent No. 6262334
; GENERAL INFORMATION:
; APPLICANT: ENDEGE, WILSON O., ET AL.
; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
; FILE REFERENCE: CCNDA-260XX
; CURRENT APPLICATION NUMBER: US/09/385,982
; EARLIER FILING DATE: 1999-08-30
; EARLIER APPLICATION NUMBER: 09/328,111
; EARLIER FILING DATE: 1999-06-08
; EARLIER APPLICATION NUMBER: 60/117,393
; EARLIER FILING DATE: 1999-01-27
; EARLIER APPLICATION NUMBER: 60/098,639
; EARLIER FILING DATE: 1998-08-31
; NUMBER OF SEQ ID NOS: 344
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 483
; LENGTH: 605
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1)...(605)
; OTHER INFORMATION: n = A,T,C or G
US-09-385-982-483

Query Match      18.4%   Score 187.2; DB 3; Length 605;
Best Local Similarity 65.5%; Pred. No. 3.2e-52;
Matches 348; Conservative 0; Mismatches 158; Indels 25; Gaps 5;

OY 485 GTACATGAGTATTACATGATTAAGTCAACGGGTAACGCTTTAAGCGTAATT 544
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 597 GGACACGANNANNAAGTCCCGGGTAACCTTAAGGTAATTGTCGAATACGTA 538
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 545 TCTGTCAACAGTAG-ATGACAATATGCCGANTATTCAGTGTCTCTGAGACCCCT 603
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 537 GATTAANAAANNAAGTTGTTGTAACAGNMAATGTTTCTTANGCTTTTCTTTTCTT 478
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 604 TCCCTCTGCTGTCCTCCAGATGGGGGTTGAGTCATATTAAATGCGCATCTCA 663
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 477 AACACCTGCCATGCTCCCAATGGCGCATTAATTCATCTTAACNGTTGTTCTGT 418
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 664 CAGTCTACTATAGCAAGTGTCTTCTTATAGACCCCTCTTAACGCAATATGCT 723
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 417 TATGCTACTATAGTATGCTTCTTCTTATGAAACCCCTTGTGATGAGCAATATGCT 358
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 724 GACCTGTACTATTAAGATCTTCTTGATATATGCAATTCGAGATTTTTCGTAGATAGTAGA 783
```


QY	872	GGGCTTTCACAGTCTTTAAACCTATTTTAAATATTCGAAACACACCTGCTTCCACAGAGTAC	931
Db	1363	YY	1422
QY	932	CAACACGTGATCGATTTGATGCGCGCCGCCCTCTAGACT	970
Db	1423	YYYYYYYYYYYYGGTACCAAAATTCCTTCTATCTCTTTAACT	1461

RESULT 4
 US-08-181-271A-1/c
 : Sequence 1, Application US/08181271A
 Patent No. 5614395
 : GENERAL INFORMATION :
 APPLICANT: Ryals, John A.
 APPLICANT: Alexander, Danny C.
 APPLICANT: Beck, James J.
 APPLICANT: Duesing, John H.
 APPLICANT: Friedrich, Leslie B.
 APPLICANT: Goodman, Robert M.
 APPLICANT: Harms, Christian
 APPLICANT: Melns, Jr., Frederick
 APPLICANT: Montoya, Alice
 APPLICANT: Moyer, Mary B.
 APPLICANT: Neuhaus, Jean-Marc
 APPLICANT: Payne, George B.
 APPLICANT: Sperison, Christoph
 APPLICANT: Stinson, Jeffrey R.
 APPLICANT: Uknes, Scott J.
 APPLICANT: Ward, Eric R.
 APPLICANT: Williams, Shericca C.
 TITLE OF INVENTION: CHEMICALLY REGULATABLE AND ANTI-PATHOGENIC
 TITLE OF INVENTION: DNA SEQUENCES AND USES THEREOF
 NUMBER OF SEQUENCES: 106
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: CIBA-GEIGY Corporation
 STREET: 7 Skyline Drive
 CITY: Hawthorne
 STATE: New York
 COUNTRY: USA
 ZIP: 10532
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/181,271A
 FILING DATE: 13-JAN-94
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/093,301
 FILING DATE: 16-JUL-1993
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/937,197
 FILING DATE: 6-NOV-1992
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/678,378
 FILING DATE: 1-APR-1991
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/305,566
 FILING DATE: 6-FEB-1989
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/165,667
 FILING DATE: 8-MAR-1988
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/042,847
 FILING DATE: 6-APR-1993
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/632,441
 FILING DATE: 21-DEC-1990
 PRIOR APPLICATION DATA:

```

1  APPLICATION NUMBER:  US 07/425,504
2  FILING DATE:  20-OCT-1989
3  PRIOR APPLICATION DATA:
4  APPLICATION NUMBER:  US 07/648,506
5  FILING DATE:  6-MAR-1992
6  PRIOR APPLICATION DATA:
7  APPLICATION NUMBER:  US 07/768,122
8  FILING DATE:  27-SEP-1991
9  PRIOR APPLICATION DATA:
10 APPLICATION NUMBER:  US 07/580,431
11 FILING DATE:  7-SEP-1990
12 PRIOR APPLICATION DATA:
13 APPLICATION NUMBER:  US 07/368,672
14 FILING DATE:  20-JUN-1989
15 PRIOR APPLICATION DATA:
16 APPLICATION NUMBER:  US 07/329,018
17 FILING DATE:  24-MAR-1989
18 PRIOR APPLICATION DATA:
19 APPLICATION NUMBER:  US 08/045,957
20 FILING DATE:  12-APR-1993
21 ATTORNEY/AGENT INFORMATION:
22 NAME:  Elmer, James Scott
23 REGISTRATION NUMBER:  36,129
24 REFERENCE/DOCKET NUMBER:  S-19825/P1/CGC 1727
25 TELECOMMUNICATION INFORMATION:
26 TELEPHONE:  (919)541-8614
27 TELEFAX:  (919)541-8689
28 INFORMATION FOR SEQ ID NO.:  1:
29     SEQUENCE CHARACTERISTICS:
30         LENGTH:  2038 base pairs
31         TYPE:  nucleic acid
32         STRANDEDNESS:  single
33         TOPOLOGY:  linear
34     MOLECULE TYPE:  cdna
35     FEATURE:
36     NAME/KEY:  CDS
37     LOCATION:  932..1435
38     US-08-181-271A-1

```

RESULT 5
US-08-449-315-1/c
Sequence 1, Application US/08449315
Patent No. 5650505
GENERAL INFORMATION:
APPLICANT: Rals, John A.
APPLICANT: Alexander, Danny C.
APPLICANT: Beck, James J.
APPLICANT: Duesing, John H.
APPLICANT: Friedrich, Leslie B.
APPLICANT: Goodman, Robert M.
APPLICANT: Harms, Christian
APPLICANT: Meins, Jr., Frederick


```

1  COMPUTER:  IBM PC compatible
2  OPERATING SYSTEM:  PC-DOS/MS-DOS
3  SOFTWARE:  PatentIn Release #1.0, Version #1.235
4  CURRENT APPLICATION NUMBER:  US/08/444,803
5  APPLICATION NUMBER DATA:
6  FILING DATE:  19-MAY-1995
7  CLASSIFICATION:  536
8  PRIOR APPLICATION DATA:
9  APPLICATION NUMBER:  08/181,271
10 FILING DATE:  13-JAN-94
11 APPLICATION NUMBER:  US 08/093,301
12 FILING DATE:  16-JUL-1993
13 PRIOR APPLICATION DATA:
14 APPLICATION NUMBER:  US 07/937,197
15 FILING DATE:  6-NOV-1992
16 PRIOR APPLICATION DATA:
17 APPLICATION NUMBER:  US 07/678,378
18 FILING DATE:  1-APR-1991
19 PRIOR APPLICATION DATA:
20 APPLICATION NUMBER:  US 07/305,566
21 FILING DATE:  6-FEB-1989
22 PRIOR APPLICATION DATA:
23 APPLICATION NUMBER:  US 07/165,667
24 FILING DATE:  8-MAR-1988
25 PRIOR APPLICATION DATA:
26 APPLICATION NUMBER:  US 08/042,847
27 FILING DATE:  6-APR-1993
28 PRIOR APPLICATION DATA:
29 APPLICATION NUMBER:  US 07/632,441
30 FILING DATE:  21-DEC-1990
31 PRIOR APPLICATION DATA:
32 APPLICATION NUMBER:  US 07/425,504
33 FILING DATE:  20-OCT-1989
34 PRIOR APPLICATION DATA:
35 APPLICATION NUMBER:  US 07/848,506
36 FILING DATE:  6-MAR-1992
37 PRIOR APPLICATION DATA:
38 APPLICATION NUMBER:  US 07/768,122
39 FILING DATE:  27-SEP-1991
40 PRIOR APPLICATION DATA:
41 APPLICATION NUMBER:  US 07/580,431
42 FILING DATE:  7-SEP-1990
43 PRIOR APPLICATION DATA:
44 APPLICATION NUMBER:  US 07/368,672
45 FILING DATE:  20-JUN-1989
46 PRIOR APPLICATION DATA:
47 APPLICATION NUMBER:  US 07/329,018
48 FILING DATE:  24-MAR-1989
49 PRIOR APPLICATION DATA:
50 APPLICATION NUMBER:  US 08/045,957
51 FILING DATE:  12-APR-1993
52 ATTORNEY/AGENT INFORMATION:
53 NAME:  Elmer, James Scott
54 REGISTRATION NUMBER:  36,129
55 TELECOMMUNICATION INFORMATION:
56 TELEPHONE:  (919)541-8614
57 TELEFAX:  (919)541-8689
58 INFORMATION FOR SEQ ID NO: 1:
59 SEQUENCE CHARACTERISTICS:
60 LENGTH: 2038 base pairs
61 TYPE: nucleic acid
62 STRANDEDNESS: linear
63 TOPOLOGY: linear
64 MOLECULE TYPE: cDNA
65 FEATURE:
66 NAME/KEY:  CDS
67 LOCATION:  932..1435
68
69 US-08-444-803-1

```

Query Match	3.58;	Score 35.4;	DB 1;	Length 2038;
Best Local Similarity	51.78;	Pred. No. 0.27;		
Matches 105;	Conservative 0;	Mismatches 96;	Indels 2;	Gaps 1

Oy	630	GGGGTGAAGCCATATTTAAACGTGGGCACATCCACAGTGGTAACTTAGCAAGGCGTTT	668
Db	314	GACGTTAAATAACTCAAAATTTGGACGTAAAGAAATATTTGGATATCTTCAGTCTTTC	255
Oy	690	CTTTAGGACC--CCCTCTTAAAGACATATATGCTGACCTGTACTATTAAGATTTTCTG	747
Db	254	TCTAATTTCTCTCCCTTAATTTCCAACTGATGTACCAATGTAATCAATCAATTTCTA	199
Oy	748	ATAATGCATTCGGAGATTTTTTGGTAGATAGTAAGATGGCGTTCCTGTTTCACTTCC	807
Db	194	TTATCTCATTCCAATCAGCGCTTGACTGGAGTATAAATAATTTCTATTTTCTTATCC	135
Oy	808	TTATCTCAGCTGCATGTGCTTC	830
Db	134	AATCAGCACTTGACTTTAGATTC	112

```

1 RESULT 7
2 US-08-449-043-1/c
3 : Sequence 1, Application US/08449043
4 : Patent No. 5689044
5 :
6 : GENERAL INFORMATION:
7 :
8 : APPLICANT: Ryals, John A.
9 : APPLICANT: Alexander, Danny C.
10 : APPLICANT: Beck, James J.
11 : APPLICANT: Duesing, John H.
12 : APPLICANT: Friedrich, Leslie B.
13 : APPLICANT: Goodman, Robert M.
14 : APPLICANT: Harms, Christian
15 : APPLICANT: Meins, Jr., Frederick
16 : APPLICANT: Montoya, Alice
17 : APPLICANT: Moyer, Mary B.
18 : APPLICANT: Neuhaus, Jean-Marc
19 : APPLICANT: Payne, George B.
20 : APPLICANT: Sperison, Christoph
21 : APPLICANT: Stinson, Jeffrey R.
22 : APPLICANT: Uknes, Scott J.
23 : APPLICANT: Ward, Eric R.
24 : APPLICANT: Williams, Sherlica C.
25 :
26 : TITLE OF INVENTION: CHEMICALLY REGULATABLE AND ANTI-PATHOGENIC
27 : TITLE OF INVENTION: DNA SEQUENCES AND USES THEREOF
28 :
29 : NUMBER OF SEQUENCES: 106
30 :
31 : CORRESPONDENCE ADDRESSES:
32 :
33 : ADDRESSEE: CIBA-GEIGY Corporation
34 : STREET: 7 Skyline Drive
35 : CITY: Hawthorne
36 : STATE: New York
37 :
38 : COUNTRY: USA
39 :
40 : ZIP: 10532
41 :
42 : COMPUTER READABLE FORM:
43 :
44 : MEDIUM TYPE: Floppy disk
45 :
46 : COMPUTER: IBM PC compatible
47 :
48 : OPERATING SYSTEM: PC-DOS/MS-DOS
49 :
50 : SOFTWARE: Patentln Release #1.0, Version #1.25
51 :
52 : CURRENT APPLICATION DATA:
53 :
54 : APPLICATION NUMBER: US/08/449,043
55 :
56 : FILING DATE: 24-MAY-1995
57 :
58 : CLASSIFICATION: 800
59 :
60 : PRIOR APPLICATION DATA:
61 :
62 : APPLICATION NUMBER: 08/181,271
63 :
64 : FILING DATE: 13-JAN-94
65 :
66 : APPLICATION NUMBER: US 08/093,301
67 :
68 : FILING DATE: 16-JUL-1993
69 :
70 : PRIOR APPLICATION DATA:
71 :
72 : APPLICATION NUMBER: US 07/937,197
73 :
74 : FILING DATE: 6-NOV-1992
75 :
76 : PRIOR APPLICATION DATA:
77 :
78 : APPLICATION NUMBER: US 07/678,378
79 :
80 : FILING DATE: 1-APR-1991
81 :
82 : PRIOR APPLICATION DATA:
83 :
84 : APPLICATION NUMBER: US 07/305,566
85 :
86 : FILING DATE: 6-FEB-1989
87 :

```



```

? ZIP: 10532
?
? COMPUTER READABLE FORM:
?
? MEDIUM TYPE: Floppy disk
?
? COMPUTER: IBM PC compatible
?
? OPERATING SYSTEM: PC-DOS/MS-DOS

```

630 GCGTTGAGTCCATATTTAACTGGCCATCCTCACAGTTGCTAACTTAGCAAGTGTCTT 689

Db 314 GACGTTAAATTAATTAATTTGACGTAAGAATTTTGGATATTTCTAGTTCTTTC 255
QY 690 CTTAGAGACC--CCCTTCTTAACGAGCAATATGTCTACCGTACTTAAGATCTTCTG 747
Db 254 TCTAATTTCCGTCCTTAATTTCCAACTGTGATGTACGATGCTATCACTTTCTA 195
QY 748 ATAAATGATTCGGAGATTTTGTGTAGATAGTGAAGTGGCTTCTGTTTACCTTCC 807
Db 194 TTATCTATTTCCAACTCAACGCGTGTGAGATTAATAATTTCTATTTCTTATCCC 135
QY 808 TTACTCAGCTGACTAGTGTTC 830
Db 134 AATCAGCACTTGACTTTAGATTTC 112

RESULT 10
US-08-455-244-1/c
; Sequence 1, Application US/08455244
; Patent No. 5789214
; GENERAL INFORMATION:
; APPLICANT: Ryals, John A.
; APPLICANT: Alexander, Danny C.
; APPLICANT: Beck, James J.
; APPLICANT: Duesing, John H.
; APPLICANT: Friedrich, Leslie B.
; APPLICANT: Goodman, Robert M.
; APPLICANT: Harms, Christian
; APPLICANT: Helms, Jr., Frederick
; APPLICANT: Montoya, Alice
; APPLICANT: Moyet, Mary B.
; APPLICANT: Neuhaus, Jean-Marc
; APPLICANT: Payne, George B.
; APPLICANT: Sperison, Christoph
; APPLICANT: Stinson, Jeffrey R.
; APPLICANT: Uknes, Scott J.
; APPLICANT: Ward, Eric R.
; APPLICANT: Williams, Shericea C.
; TITLE OF INVENTION: CHEMICALLY REGULATABLE AND ANTI-PATHOGENIC
; TITLE OF INVENTION: DNA SEQUENCES AND USES THEREOF
; NUMBER OF SEQUENCES: 106
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CIBA-GEIGY Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: New York
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/455,244
; FILING DATE: 31-MAY-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/181,271
; FILING DATE: 13-JAN-94
; APPLICATION NUMBER: US 08/093,301
; FILING DATE: 16-JUL-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/937,197
; FILING DATE: 6-NOV-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/678,378
; FILING DATE: 1-APR-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/305,566
; FILING DATE: 6-FEB-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/165,667

; FILING DATE: 8-MAR-1988
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/042,847
; FILING DATE: 6-APR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/632,441
; FILING DATE: 21-DEC-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/425,504
; FILING DATE: 20-OCT-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/848,506
; FILING DATE: 6-MAR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/768,122
; FILING DATE: 27-SEP-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/580,431
; FILING DATE: 7-SEP-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/368,672
; FILING DATE: 20-JUN-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/329,018
; FILING DATE: 24-MAR-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/045,957
; FILING DATE: 12-APR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Elmer, James Scott
; REGISTRATION NUMBER: 36,129
; REFERENCE/DOCKET NUMBER: S-19825/P1/CGC 1727
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (919)541-8614
; TELEFAX: (919)541-8689
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2038 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 932..1435
; US-08-455-244-1

Query Match 3.5%; Score 35.4; DB 1; Length 2038;
Best Local Similarity 51.7%; Pred. No. 0.27; Mismatches 96; Indels 2; Gaps 1;
Matches 105; Conservative 0;

QY 630 GCGGTGAGTCATATTTAACTGCCATCTCACAGTTGCTAAGCAAGTCTTTT 689
Db 314 GACGTTAAATTAATTAATTTGACGTAAGAATTTGGATATTTCTAGTTCTTTC 255
QY 690 CTTAGAGACC--CCCTTCTTAACGAGCAATATGTCTACCGTACTTAAGATCTTCTG 747
Db 254 TCTAATTTCCGTCCTTAATTTCCAACTGTGATGTACGATGCTATCACTTTCTA 195
QY 748 ATAAATGATTCGGAGATTTTGTGTAGATAGTGAAGTGGCTTCTGTTTACCTTCC 807
Db 194 TTATCTATTTCCAACTCAACGCGTGTGAGATTAATAATTTCTATTTCTTATCCC 135
QY 808 TTACTCAGCTGACTAGTGTTC 830
Db 134 AATCAGCACTTGACTTTAGATTTC 112

RESULT 11
US-08-454-876-1/c
; Sequence 1, Application US/08454876
; Patent No. 5804693
; GENERAL INFORMATION:


```

      FILING DATE: 20-JUN-1989
      PRIOR APPLICATION DATA:
      APPLICATION NUMBER: US 07/329,018
      FILING DATE: 24-MAR-1989
      PRIOR APPLICATION DATA:
      APPLICATION NUMBER: US 08/045,957
      FILING DATE: 12-APR-1993
      ATTORNEY/AGENT INFORMATION:
      NAME: Elmer, James Scott
      REGISTRATION NUMBER: 36,129
      REFERENCE/DOCKET NUMBER: S-19825/P1/GC 1727
      TELECOMMUNICATION INFORMATION:
      TELEPHONE: (919)541-8614
      TELEFAX: (919)541-8689
      INFORMATION FOR SEQ ID NO: 1:
      SEQUENCE CHARACTERISTICS:
      LENGTH: 2038 base pairs
      TYPE: nucleic acid
      STRANDEDNESS: single
      TOPOLOGY: linear
      MOLECULE TYPE: cDNA
      FEATURE:
      NAME/KEY: CDS
      LOCATION: 932..1435
      US-08-454-876-1

Query Match          3.5%; Score 35.4; DB 1; Length 2038;
Best Local Similarity 51.7%; Pred. No. 0.27; 96; Indels 2; Gaps 1
Matches 105; Conservative 0; Mismatches

QY 630 GGCCTGAGTCATATTTAAATCGCCATCCCTCACAGTTGCTAACTTAGCAAGTCTTTT 689
   ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 314 GACGTTAATTAATACATCAATTAATTTGAGCTAAAGAATTAATTTGGATTTCTTGACGTTCTTC 255
QY 690 CTTTAGAGACC--CCCTTCTTAACGACGAAATGTCTGACCGTACTATAAGATCTTTCTG 747
   ||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 254 TCTAATTTCTCTCCCTTAATTTCCAACTTGATGTTACTCAAGTACTATCAATCACTTTCTTA 195
QY 748 AATAATGCAATCGAGATTTTGTGTAGATAGTAGAAGTGGCTTCTGTTTTCACCTTCC 807
   ||||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 194 TTATCTCATTCATCCATCAACGCGTTGATCGAGTAAATTAATTTCTATTTTCTTAGTCC 135
QY 808 TTTACTAGCTGACTAGTGTCTC 830
   | | ||||| ||| |||
Db 134 AATCAGCACTTGACTTAGATTTC 112
   ||| ||||| ||| |||

RESULT 12
US-08-457-364-1/C
: Sequence 1, Application US/08457364
: Patent No. 5847258
: GENERAL INFORMATION:
: APPLICANT: Ryals, John A.
: APPLICANT: Alexander, Danny C.
: APPLICANT: Beck, James J.
: APPLICANT: Duesing, John H.
: APPLICANT: Friedrich, Leslie B.
: APPLICANT: Goodman, Robert M.
: APPLICANT: Harms, Christian
: APPLICANT: Melns, Jr., Frederick
: APPLICANT: Montoya, Alice
: APPLICANT: Moyer, Mary B.
: APPLICANT: Neuhaus, Jean-Marc
: APPLICANT: Payne, George B.
: APPLICANT: Sperison, Christoph
: APPLICANT: Stinson, Jeffrey R.
: APPLICANT: Ukens, Scott J.
: APPLICANT: Ward, Eric R.
: APPLICANT: Williams, Shericea C.
: TITLE OF INVENTION: CHEMICALLY REGULATABLE AND ANTI-PATHOGENIC
: NUMBER OF SEQUENCES: 106
: CORRESPONDENCE ADDRESS: 106

```



```

1 APPLICATION NUMBER: US 07/937,197
2 FILING DATE: 6-NOV-1992
3 PRIOR APPLICATION DATA:
4 APPLICATION NUMBER: US 07/678,378
5 FILING DATE: 1-APR-1991
6 PRIOR APPLICATION DATA:
7 APPLICATION NUMBER: US 07/305,566
8 FILING DATE: 6-FEB-1989
9 PRIOR APPLICATION DATA:
10 APPLICATION NUMBER: US 07/165,667
11 FILING DATE: 8-MAR-1988
12 PRIOR APPLICATION DATA:
13 APPLICATION NUMBER: US 08/042,847
14 FILING DATE: 6-APR-1993
15 PRIOR APPLICATION DATA:
16 APPLICATION NUMBER: US 07/632,441
17 FILING DATE: 21-DEC-1990
18 PRIOR APPLICATION DATA:
19 APPLICATION NUMBER: US 07/425,504
20 FILING DATE: 20-OCT-1989
21 PRIOR APPLICATION DATA:
22 APPLICATION NUMBER: US 07/848,506
23 FILING DATE: 6-MAR-1992
24 PRIOR APPLICATION DATA:
25 APPLICATION NUMBER: US 07/768,122
26 FILING DATE: 27-SEP-1991
27 PRIOR APPLICATION DATA:
28 APPLICATION NUMBER: US 07/580,431
29 FILING DATE: 7-SEP-1990
30 PRIOR APPLICATION DATA:
31 APPLICATION NUMBER: US 07/368,672
32 FILING DATE: 20-JUN-1989
33 PRIOR APPLICATION DATA:
34 APPLICATION NUMBER: US 07/329,018
35 FILING DATE: 24-MAR-1989
36 PRIOR APPLICATION DATA:
37 APPLICATION NUMBER: US 08/045,957
38 FILING DATE: 12-APR-1993
39 ATTORNEY/AGENT INFORMATION:
40 NAME: Elmer, James Scott
41 REGISTRATION NUMBER: 36,129
42 REFERENCE/DOCKET NUMBER: S-19825/P1/GCG 1727
43 TELECOMMUNICATION INFORMATION:
44 TELEPHONE: (919)541-8614
45 TELEFAX: (919)541-8689
46 INFORMATION FOR SEQ ID NO: 1:
47 SEQUENCE CHARACTERISTICS:
48 LENGTH: 2038 base pairs
49 TYPE: nucleic acid
50 STRANDEDNESS: single
51 TOPOLOGY: linear
52 MOLECULE TYPE: cdna
53 FEATURE:
54 NAME/KEY: CDS
55 LOCATION: 932..1435
56 US-08-456-262-1

```

[illegible]

Qy	808	TTTACTCAGCTGACTAGTGC	830
Db	134	AATCACGACTTGACTTTAGATTC	112

RESULT 14
US-08-456-240-1/c
Sequence 1, Application US/08456240
Patent No. 5856154

GENERAL INFORMATION:

APPLICANT: Ryals, John A.
APPLICANT: Alexander, Danny C.
APPLICANT: Beck, James J.
APPLICANT: Duesing, John H.
APPLICANT: Friedrich, Leslie B.
APPLICANT: Goodman, Robert M.
APPLICANT: Harms, Christian
APPLICANT: Melns, Jr., Frederick
APPLICANT: Montoya, Alice
APPLICANT: Meyer, Jean B.
APPLICANT: Neuhaus, Jean-Marc
APPLICANT: Payne, George B.
APPLICANT: Sperlson, Christoph
APPLICANT: Stinson, Jeffrey R.
APPLICANT: Uknes, Scott J.
APPLICANT: Ward, Eric R.
APPLICANT: Williams, Sherlica C.

TITLE OF INVENTION: CHEMICALLY REGULATABLE AND ANTI-PATHOGENIC
TITLE OF INVENTION: DNA SEQUENCES AND USES THEREOF

NUMBER OF SEQUENCES: 106

CORRESPONDENCE ADDRESS:

ADDRESSEE: CIBA-GEIGY Corporation
STREET: 7 Skyline Drive
CITY: Hawthorne
STATE: New York
COUNTRY: USA
ZIP: 10532

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/456,240
FILING DATE: 31-MAY-1995

CLASSIFICATION: 800

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/181,271
FILING DATE: 13-JAN-94
APPLICATION NUMBER: US 08/093,301
FILING DATE: 16-JUL-1993

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/937,197
FILING DATE: 6-NOV-1992

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/678,378
FILING DATE: 1-APR-1991

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/305,566
FILING DATE: 6-FEB-1989

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/165,667
FILING DATE: 8-MAR-1988

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/042,847
FILING DATE: 6-APR-1993

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/632,441
FILING DATE: 21-DEC-1990

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/425,504
FILING DATE: 20-OCT-1989

PRIOR APPLICATION DATA:

```

? APPLICATION NUMBER: US 07/848,506
? FILING DATE: 6-MAR-1992
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 07/768,122
? FILING DATE: 27-SEP-1991
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 07/580,431
? FILING DATE: 7-SEP-1990
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 07/368,672
? FILING DATE: 20-JUN-1989
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 07/329,018
? FILING DATE: 24-MAR-1989
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 08/045,957
? FILING DATE: 12-APR-1993
? ATTORNEY/AGENT INFORMATION:
? NAME: Elmer, James Scott
? REGISTRATION NUMBER: 36,129
? REFERENCE/DOCKET NUMBER: S-19825/P1/CSC 1727
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: (919)541-8614
? TELEFAX: (919)541-8689
? INFORMATION FOR SEQ ID NO: 1:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 2038 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? MOLECULE TYPE: cDNA
? FEATURE:
? NAME/KEY: CDS
? LOCATION: 932..1435
US-08-456-240-1

Query Match          3.5% Score 35.4; DB 2; Length 2038;
Best Local Similarity 51.7%; Pred No. 0.27;
Matches 105; Conservativity 0; Mismatches 96; Indels 2; Gaps 1

OY      630   GCGGTGAGTCGCATATTAAACCTGCCATCCTCACAGATTGCTAAGTAGCAAGTCTTT    689
           ||||| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db       314   GACGTTAAATAACTCATCAAAATTTGGACGTAAGAATAATTTGGATATCTTCAGTCTTT    255
OY      690   CTTAGAGACC--CCCTCTTAAACGAGCAANTGCTGACCGTACTATAAGATCTTGCG    747
           ||||| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db       254   TCTAATTTCCCTCCCTTAATTTCCACAGCTTGATGTAAGTACTATCATCAACTTTTGA    195
OY      748   ATAAATGCAATTCGAGACTTTTTTTGGTAGATAGTAGAAGTGCCTCCCTTTACACTTCC    807
           ||||| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db       194   TTATGCTATTCACAATCAGCGGTGATGTCGAGATATAAATAATTCTAATTTTCTATGCC    135
OY      808   TTTACTCAGCTGACTAGTGCCTC    830
           | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db       134   AATCAGACTTGACTTTAGATTTC    112

RESULT 15
US-08-455-736-1/c
? Sequence 1, Application US/08455736
? Patent No. 5880328
? GENERAL INFORMATION:
? APPLICANT: Ryals, John A.
? APPLICANT: Alexander, Danny C.
? APPLICANT: Beck, James J.
? APPLICANT: Duesing, John H.
? APPLICANT: Friedrich, Leslie B.
? APPLICANT: Goodman, Robert M.
? APPLICANT: Harms, Christian
? APPLICANT: Meins, Jr., Frederick
? APPLICANT: Montoya, Alice
? APPLICANT: Moyer, Mary B.
? APPLICANT: Neuhaus, Jean-Marc

```

APPLICANT: Payne, George B
 APPLICANT: Sperison, Christoph
 APPLICANT: Stinson, Jeffrey R.
 APPLICANT: Uknes, Scott J.
 APPLICANT: Ward, Eric R.
 APPLICANT: Williams, Shericca C.
 TITLE OF INVENTION: CHEMICALS REGULATABLE AND ANTI-PATHOGENIC
 TITLE OF INVENTION: DNA SEQUENCES AND USBS THEREOF
 NUMBER OF SEQUENCES: 106
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: CIBA-GEIGY Corporation
 STREET: 7 Skyline Drive
 CITY: Hawthorne
 STATE: New York
 COUNTRY: USA
 ZIP: 10532
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/455,736
 FILING DATE: 31-MAY-1995
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/181,271
 FILING DATE: 13-JAN-1994
 APPLICATION NUMBER: US 08/093,301
 FILING DATE: 16-JUL-1993
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/937,197
 FILING DATE: 6-NOV-1992
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/678,378
 FILING DATE: 1-APR-1991
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/305,566
 FILING DATE: 6-FEB-1989
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/165,667
 FILING DATE: 8-MAR-1988
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/042,847
 FILING DATE: 6-APR-1993
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/632,441
 FILING DATE: 21-DEC-1990
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/425,504
 FILING DATE: 20-OCT 1989
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/848,506
 FILING DATE: 6-MAR-1992
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/768,122
 FILING DATE: 27-SEP-1991
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/580,431
 FILING DATE: 7-SEP-1990
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/368,672
 FILING DATE: 20-JUN-1989
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/329,018
 FILING DATE: 24-MAR-1989
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/045,957
 FILING DATE: 12-APR-1993
 ATTORNEY/AGENT INFORMATION:
 NAME: Elmer, James Scott
 REGISTRATION NUMBER: 36,129
 REFERENCE/DOCKET NUMBER: S-19825/P1/CGC 1727

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 25, 2003, 22:04:33 ; Search time 248.548 seconds
(without alignments)
9198.941 Million cell updates/sec

Title: US-09-717-321A-15

Perfect score: 1017
Sequence: 1 cccctattctgtctcagatf.....ccttgggtctgtgagttc 1017

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1517243 seqs, 1124081882 residues

Total number of hits satisfying chosen parameters: 3034486

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Published Applications_NA:*

- 1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/2/pubpna/PCR_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/2/pubpna/CTOS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
- 10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
- 12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
- 13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
- 14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
- 15: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
- 16: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
- 17: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match length	ID	Description
1	290.6	28.6	1644 14 US-10-153-668-48	Sequence 48, Appl
2	290.6	28.6	1646 14 US-10-153-668-46	Sequence 46, Appl
3	290.6	28.6	2051 9 US-09-923-302-255	Sequence 255, App
4	290.6	28.6	2941 14 US-10-198-846-9874	Sequence 9874, Ap
5	285.2	28.0	1232 10 US-09-880-107-1632	Sequence 1632, Ap
6	283.2	27.8	2313 14 US-10-198-846-12919	Sequence 12919, A
7	195.2	19.0	594 14 US-10-066-543-2864	Sequence 2864, Ap
8	193.2	19.0	602 10 US-09-878-178-1654	Sequence 1654, Ap
9	193.2	19.0	602 13 US-10-046-935-1654	Sequence 1654, Ap
10	193.2	19.0	602 14 US-10-146-502-1654	Sequence 1654, Ap
11	187.8	18.5	549 14 US-10-066-543-2878	Sequence 2878, Ap
12	187.8	18.5	567 14 US-10-066-543-2964	Sequence 2964, Ap
13	187.2	18.4	605 11 US-09-871-161-483	Sequence 483, App
14	185.4	18.2	477 14 US-10-066-543-3022	Sequence 3022, App
15	172.2	16.9	432 10 US-09-960-352-7612	Sequence 7612, Ap
16	164	16.1	2167 12 US-10-101-510-500	Sequence 500, App

17	147.4	14.5	409 14 US-10-066-543-13	Sequence 13, Appl
18	126.4	12.4	285 9 US-09-294-093B-506	Sequence 506, App
19	123.2	12.1	416 10 US-09-960-352-8781	Sequence 8781, Ap
20	108.8	10.7	293 14 US-10-066-543-3077	Sequence 3077, Ap
21	103.2	10.1	3740 11 US-09-764-891-9984	Sequence 9984, Ap
22	98.2	9.7	581 10 US-09-998-598-972	Sequence 972, App
23	91.6	9.0	565 10 US-09-998-598-966	Sequence 296, App
24	85.6	8.4	492 10 US-09-878-178-1503	Sequence 1503, App
25	85.6	8.4	492 13 US-10-046-935-1503	Sequence 1503, Ap
26	85.6	8.4	492 14 US-10-046-935-1503	Sequence 1503, Ap
27	82.2	8.1	398 11 US-09-918-995-34569	Sequence 34569, A
28	68	6.7	486 11 US-09-871-161-133	Sequence 133, App
29	62.2	6.1	426 10 US-09-960-352-7342	Sequence 7342, Ap
30	59	5.8	404 10 US-09-960-352-14206	Sequence 14206, A
31	57.6	5.7	1651 14 US-10-198-846-13019	Sequence 13019, A
32	52.4	5.2	162 10 US-09-920-100A-1549	Sequence 1549, Ap
33	52.4	5.2	162 13 US-10-033-528-1549	Sequence 1549, Ap
34	40.8	4.0	269 9 US-09-923-876-6297	Sequence 6297, Ap
35	37.4	3.7	565 13 US-10-311-455-1504	Sequence 1504, Ap
36	37.2	3.7	565 12 US-10-027-632-133365	Sequence 133365, Ap
37	37	3.6	260209 12 US-10-025-966A-23	Sequence 23, Appl
38	37	3.6	260209 12 US-10-265-071-23	Sequence 23, Appl
39	36.6	3.6	1581 14 US-10-198-846-6480	Sequence 6480, Ap
40	36.4	3.6	725 13 US-10-027-632-11287	Sequence 11287, A
41	36.2	3.6	968 13 US-10-027-632-121396	Sequence 121396, A
42	36	3.5	5098 14 US-10-192-867-1	Sequence 1, Appl1
43	35.2	3.5	560 13 US-10-027-632-268476	Sequence 268476, Ap
44	34.8	3.4	454 13 US-10-027-632-182180	Sequence 182180, Ap
45	34.8	3.4	454 13 US-10-027-632-182181	Sequence 182181, Ap

ALIGNMENTS

RESULT 1
US-10-153-668-48
Sequence 48, Application US/10153668
Publication No. US20030092616A1
GENERAL INFORMATION:
APPLICANT: HONDA, Goichi
APPLICANT: MATSUDA, AKIO
APPLICANT: MORIMATSU, Shuji
TITLE OF INVENTION: ISHIZAWA, Kenya
FILE REFERENCE: 1254-0207P
CURRENT APPLICATION NUMBER: US/10/153,668
CURRENT FILING DATE: 2002-05-24
PRIOR APPLICATION NUMBER: US 60/293,172
PRIOR FILING DATE: 2001-05-25
PRIOR APPLICATION NUMBER: US 60/316,031
PRIOR FILING DATE: 2001-08-31
PRIOR APPLICATION NUMBER: US 60/328,403
PRIOR FILING DATE: 2001-10-12
PRIOR APPLICATION NUMBER: JP 2001-157043
PRIOR FILING DATE: 2001-05-25
PRIOR APPLICATION NUMBER: JP 2001-260681
PRIOR FILING DATE: 2001-08-30
PRIOR APPLICATION NUMBER: JP 2001-313175
PRIOR FILING DATE: 2001-10-10
NUMBER OF SEQ ID NOS: 488
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 48
LENGTH: 1644
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (109)..(447)
US-10-153-668-48
Query Match 28.6%; Score 290.6; DB 14; Length 1644;
Best Local Similarity 64.5%; Pred. No. 2.5e-83;
Matches 664; Conservative 0; Mismatches 304; Indels 61; Gaps 13;


```

OY      929  TACCAACACGTCTATGATGATGCCGCCCTCTAGACTACACCGAGACATG 988
        TACCAACACGTCTATGATGATGCCGCCCTCTAGACTACACCGAGACATG 988
Db      1687 TACCGACACGTCTACTCTGACCAATACAGT-ACCCTCTTACTCCGCCACGGACACAGC 1745
OY      989 CTTCCGGTA 997
        TTTT
Db      1746 CCTCCTGTA 1754

RESULT 4
US-10-198-846-9874
: Sequence 9874, Application US/10198846
: Publication No. US20030099974A1
: GENERAL INFORMATION:
: APPLICANT: Liilie, James
: APPLICANT: Xu, Yongyao
: APPLICANT: Wang, Youzhen
: APPLICANT: Steinmann, Kathleen
: TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS
: TITLE OF INVENTION: FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
: FILE REFERENCE: MRI-049
: CURRENT APPLICATION NUMBER: US/10/198,846
: PRIOR FILING DATE: 2002-07-18
: PRIOR FILING DATE: 2001-07-18
: NUMBER OF SEQ ID NOS: 14084
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 9874
: LENGTH: 2941
: TYPE: DNA
: ORGANISM: Homo sapiens
US-10-198-846-9874

```

Query Match	28.6%	Score	290.6	DB	14	Length	2941
Best Local Similarity	64.5%	Pred. No.	3.7e-83				
Matches	664	Conservative	0	Mismatches	304	Indels	61
						Gaps	13

OY	1	CCCCATATTCCTGGCAGATTAAAGAAATGGCCAAATACCTGTGAACAAATGGC----	-GTT	57
Db	1195	CCCCCATTCCTGGTCAGATTAAAGATTGCCAAATACCTGTGAACAAATGGCATTGTT	111	111
OY	58	GTCGTGAGAACAGCTAAGCACTAAGCTGTGTGAGACATTGTCTTAAAGAAAGCTGACG	117	
Db	1255	GTTCGGGCTGAGG--GGTGCAGACCCCTCCGTAAGC-----TCCAGACGGTGTGACACA	169	
OY	118	TTTCGGGCTGAGG--GGTGCAGACCCCTCCGTAAGC-----TCCAGACGGTGTGACACA	169	
Db	1315	TTTCGGGCTGAGG--GGTGCAGACCCCTCCGTAAGC-----TCCAGACGGTGTGACACA	169	
OY	170	GCACAGCCTCCCTTAATGACACGACCTGCCATGTAAAGCAGCAGCTTAATCTTACGCCCATAGCT	229	
Db	1375	GACACAGCCTCCCGAATGAAAGGTTGCCATTGACATCAACAGTGAATTGACAGCAGTGT	143	
OY	230	CATTACGTAATCTTGTACTGTACTGACGTCAAGATGGGTGTAAAGCTCTGCTCTTTGATTTC	289	
Db	1435	CCCAGACATTAACATTGTACTGTAAATGGAATGAGCTGAGCAGCTCAGCTTTTGGATCACTC	1494	
OY	290	TAGTAGAGTTCCTCTAAATATACCAAGCTGACCGGCTCTTGACAGCTTGTGAACAGACACTGTGC	349	
Db	1495	TTTGTGATTTCATAGCGAGTTCCTGTGACACAGCTTTTGCGGGAGATTTTGAACAGACTG--	1555	
OY	350	TCTCTGTGTGCTCTTAACGAAGATTTCCTAGTCGTGGGCTGTGCTGCGGTGAGACTG	409	
Db	1553	----CTATTTCCTCTAATGAAGAAATTCGTG--TAGCTGTGGGTGTGCGGGGTGGGTGT	1606	
OY	410	GTGAACAACAGCGTCATCAAAAGGACAGACAGATATTTGAC--TAATATGAAGTAGAGAT	468	
Db	1607	GT-----GTGATCCAAAGGACAAAGACAGATATTTTGAACAAATACGAAATGAGAG--	1654	
OY	469	TAATTTCACTACATGTGTACATGGAGTAA--TTCAACTGAATAAAGTGTACAGGGTAA	525	

Db	1655	-ATTACACACACTTGTACNAGAAATGAAGTGTACACGGGTAAAACTGTAAGAAAGTTA	1712
QY	526	AGCTTTTAAAGGTTAATTTCTGTGCAAAACACTACATGACAAATGGCCGATCTTATCAGTG	585
Db	1713	ATTTCTGTCAAAATGACAGTAGATGATGAAGAAAGAAAGGTTGTATTTTCAGGAAATTTTTTCT	1772
QY	586	TCCTCTTGAGCCCCCTCCCTCCCTGCTGCTCCCTCCACAGTGGGGGGTTGAGTGCATAT	645
Db	1773	TAACTTTTCTCTTCTCTTACACCTGCCATGCCCTCCCCCAAATGGGCAATTTTAATTCATCT	1832
QY	646	TTAAACAGGCCATCCTCACAGTTGCTTAACATTAGCAAGTGTCTTTCTTTAGAACCCCTTC	705
Db	1833	TTAAACAGGTTGTTCTGTGTAGTGGCTAACTAGTAAGTGCTTTCTTATAGAACCCCTTC	1892
QY	706	TTAAACAGCAATATGTCTGACCTGTACTATAGATCTTCTTGATTAATGATTCGAGATTT	765
Db	1893	TGACTGTAGCAATATAGCCT-CCTTGTATTTATTAATCTTTCGATAAATGCAATTAAGAAAGTT	1951
QY	766	TTTTTGTAGATATATAGAGTGCCTTCTGTTTCACCTTCCTTACTACAGCTACAGT	825
Db	1952	TTTTTGTGCAGATTATGATAAAGTGCCTTCCAGTTAC- - -TTTATTCAGACACTAATAGT	2006
QY	826	GCTTCCCTCGTTTCTTCTAGTAACTAGGGGTATGAATACAGCTGCGGCTTACAGTTTTT	885
Db	2007	GCTTCCCTAGATTCTCTAGTAACTAGGTATGAATATCATGTGTGGACCTTATAGTTTTT	2066
QY	886	TAAACTATTTTATAGATA- - - - -TTCTGAACAATCACTGTCTTGCCAGAG	928
Db	2067	TAAATATATTTTATAGATTAATCTTAAACATAGAACCTTCTTAAATCAATCACTGTCTTGCCAGAT	2126
QY	929	TACCAACACTGTCAATGTGATTTGATGATCGGCCCCCTCTGTAGACTCAACCCACGGGACATG	988
Db	2127	TACCGACACTGTCACTGTGACCAATATATG-ACCTCTTTTACTCTCCGCCACGGCGAGACACAG	2185
QY	989	CTTCCGGTA 997	
Db	2186	CTCTCTGTA 2194	

Db 2186 CCTCCTGTA 2194

```

RESULT 5
US-09-880-107-1632
; Sequence 1632, Application US/09880107
; Patent No. US20020142981A1
; GENERAL INFORMATION:
; APPLICANT: Horne, Darci T.
; APPLICANT: Vockley, Joseph G.
; APPLICANT: Scherf, Uwe
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
; FILE REFERENCE: 44921-5028-NO
; CURRENT APPLICATION NUMBER: US/09/880,107
; CURRENT FILING DATE: 2001-06-14
; PRIOR APPLICATION NUMBER: US 60/211,379
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: US 60/237,054
; PRIOR FILING DATE: 2000-10-02
; NUMBER OF SEQ ID NOS: 3950
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1632
; LENGTH: 1232
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20020142981A1 D25274
US-09-880-107-1632

```

Query Match	28.0%	Score 285.2	DB 10	length 1232
Best Local Similarity	64.4%	Pred. No. 1.1e-81		
Matches 658	Conservative	0	Mismatches 303	Indels 61
				Gaps 13

07 TCTTGGCTAGATTTAGAAATTTGCCAAAATACCTGTGAACCTAAGTTGC---GTTTGGCTGA 64

1 TCTTGGCTAGATTTAGAAATTTGCCAAAATACCTGTGAACCTAAGTTGC---GTTTGGCTGA 60


```
OY 709 ACAGCAATATGCTGACCTGACTATTAAGATCTTTCTGATAATGCAATTCGGAGATTTT 768
    |||||
DB 1923 CTGAGCAATATGCTT-CTTGATTTATTAATCTTTCTGATAATGCAATTAAGAGTTT 1981
    |||||
OY 769 TTGGTAGATAGTAGAAGTGGCTGCTGTTTACACTCCCTTTCTGACGAGTACTAGTGGT 828
    |||||
DB 1982 TTGTCAATTAAGTAAGAGTGGCTTTCCATGTTAC-----TTTATTCAGAGCTTAATAGTGT 2036
    |||||
OY 829 TCCCTGCTTTTCTAGTAACGTGGGTAGAAATCAGCTGCTGGCGCTTTACAGTTTAA 888
    |||||
DB 2037 TTCCCTAGTTTCTAGTAACGTGGTAATAATCAGTGTGGCAGCTTTATGTTTAA 2096
    |||||
OY 889 ACATTTTATGATA-----TTCTGAACATCACTGCTGCTCCAGAGTAC 931
    |||||
DB 2097 AATATTTTATGATAATCTTAATACTAGTAACCTTTCTTAACATCACTGCTGCGCAGGTAC 2156
    |||||
OY 932 CAACACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 991
    |||||
DB 2157 CGACACTGCTCACTTGACATTAAGT-ACCCTCTTACTGCTGCGCAGCGACACAGCCCT 2215
    |||||
OY 992 CCGGTA 997
    |||||
DB 2216 CCGTGA 2221
    |||||
```

RESULT 7

```
US-10-066-543-2864
; Sequence 2864, Application US/10066543
; Publication No. US2003087818A1
; GENERAL INFORMATION:
; APPLICANT: Jlang, Yunglu
; APPLICANT: Pyle, Ruth A.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Indirias, Carol yoseph
; APPLICANT: Lodes, Michael J.
; APPLICANT: Secrist, Heather
; APPLICANT: Carter, Darick
; APPLICANT: Fanger, Gary R.
; APPLICANT: Smith, Carole L.
; APPLICANT: Durham, Margareta
; APPLICANT: Stolk, John A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.563
; CURRENT APPLICATION NUMBER: US/10/066,543
; CURRENT FILING DATE: 2002-01-31
; NUMBER OF SEQ ID NOS: 3417
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2864
; LENGTH: 594
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-066-543-2864
```

Query Match 19.2%; Score 195.2; DB 14; Length 594;

Best Local Similarity 70.9%; Pred. No. 1.4e-52; Indels 40; Gaps 7;

Matches 373; Conservative 0; Mismatches 113;

```
OY 505 GAATAAAAGTGTACAGGGTAAGCTTTTAAAGCTTAATTTCTGTCAAA--CAGTAGATG 562
    |||||
DB 6 GAATGAAGTGTACAGGGTAAGCTTTTAAAGCTTAATTTCTGTCAAAAGCAGTAGATG 65
    |||||
OY 563 ACAAAAT---GGCCGATCTTATCAGTGTCTCT-----CTTGAGCCCCCTTCC 608
    |||||
DB 66 ATGAAGAAAGTGTGTGATATATCAGAAATGTTTCTTAAGCTTTCTTCTCTACAC 125
    |||||
OY 609 CTGCTGCTCCCTCCAGATGGGGCTTGAAGTCAATTAATTAAGTGGCAGTCCAGAGT 668
    |||||
DB 126 CTGCTGCTCCCTCCAGATGGGGCTTGAAGTCAATTAATTAAGTGGTGTCTGTAGTC 185
    |||||
OY 669 GCTAATTAAGTCAAGTCTTTTCTTGTAGACCCCTTCTTAAGCAGCAATATGCTGACCT 728
    |||||
```

```
DB 186 GCTAATTAAGTCAAGTCTTTTCTTATAGAACCCCTTCTGACTGAGCAATATGCTT-CTT 244
    |||||
OY 729 GCTAATTAAGTCAAGTCTTTTCTTATAGAACCCCTTCTGACTGAGCAATATGCTT-CTT 788
    |||||
DB 245 GCTAATTAAGTCAAGTCTTTTCTTATAGAACCCCTTCTGACTGAGCAATATGCTT-CTT 304
    |||||
OY 789 GTTCTGCTTTTCACTCTTTTCTGAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAG 848
    |||||
DB 305 TTGCTGCTTTTCACTCTTTTCTGAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAG 359
    |||||
OY 849 TTGGTAGAATACAGCTGCTGGGTAGAAATCAGCTGCTGGCGCTTTACAGTTTAA 901
    |||||
DB 360 TAGTGTAAATAATCAATGCTGCTGAGCTTTATGTTTAAATTTTATGATTAATCTTA 419
    |||||
OY 902 -----TTCTGAACATCACTGCTGCTGCGCAGAGTACCAACAGTCAATGATTA 951
    |||||
DB 420 AACTATGAACCTTCTTAACATCACTGCTGCGCAGAGTACCAACAGTCAATGATTA 479
    |||||
OY 952 TGGCGCCCTCTGAGACCTTCAAGCAGCGGACAGCAGCTTCCGGTA 997
    |||||
DB 480 TACTG-ACCCTCTTACTGCTGCGCAGCGACAGCAGCTTCTGTA 524
    |||||
```

RESULT 8

```
US-09-878-178-1654/C
; Sequence 1654, Application US/09878178
; Patent No. US2002017752A1
; GENERAL INFORMATION:
; APPLICANT: Jlang, Yunglu
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Secrist, Heather
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.527
; CURRENT APPLICATION NUMBER: US/09/878,178
; CURRENT FILING DATE: 2001-06-08
; NUMBER OF SEQ ID NOS: 2237
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1654
; LENGTH: 602
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(602)
; OTHER INFORMATION: n = A,T,C or G
US-09-878-178-1654
```

Query Match 19.0%; Score 193.2; DB 10; Length 602;

Best Local Similarity 70.5%; Pred. No. 6.2e-52; Indels 40; Gaps 7;

Matches 371; Conservative 0; Mismatches 115;

```
OY 505 GAATAAAAGTGTACAGGGTAAGCTTTTAAAGCTTAATTTCTGTCAAA--CAGTAGATG 562
    |||||
DB 592 GAATGAAGTGTACAGGGTAAGCTTTTAAAGCTTAATTTCTGTCAAAAGCAGTAGATG 533
    |||||
OY 563 ACAAAAT---GGCCGATCTTATCAGTGTCTCT-----CTTGAGCCCCCTTCC 608
    |||||
DB 532 ATGAAGAAAGTGTGTGATATATCAGAAATGTTTCTTAAGCTTTCTTCTCTACAC 473
    |||||
OY 609 CTGCTGCTCCCTCCAGATGGGGCTTGAAGTCAATTAATTAAGTGGCAGTCCAGAGT 668
    |||||
DB 472 CTGCTGCTCCCTCCAGATGGGGCTTGAAGTCAATTAATTAAGTGGTGTCTGTAGTC 413
    |||||
OY 669 GCTAATTAAGTCAAGTCTTTTCTTGTAGACCCCTTCTTAAGCAGCAATATGCTGACCT 728
    |||||
DB 412 GCTAATTAAGTCAAGTCTTTTCTTATAGAACCCCTTCTGACTGAGCAATATGCTT-CTT 354
    |||||
OY 729 GCTAATTAAGTCAAGTCTTTTCTTATAGAACCCCTTCTGACTGAGCAATATGCTT-CTT 788
    |||||
DB 353 GCTAATTAAGTCAAGTCTTTTCTTATAGAACCCCTTCTGACTGAGCAATATGCTT-CTT 294
    |||||
OY 789 GTTCTGCTTTTCACTCTTTTCTGAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAG 848
    |||||
```

[illegible]


```
;; FILE REFERENCE: CCDNA-260XX
;; CURRENT APPLICATION NUMBER: US/09/871.161
;; CURRENT FILING DATE: 2001-05-31
;; PRIOR APPLICATION NUMBER: 09/328.111
;; PRIOR FILING DATE: 1999-06-08
;; PRIOR APPLICATION NUMBER: 60/117.393
;; PRIOR FILING DATE: 1999-01-27
;; PRIOR APPLICATION NUMBER: 60/098.639
;; PRIOR FILING DATE: 1998-08-31
;; NUMBER OF SEQ ID NOS: 544
;; SOFTWARE: FastSeq for Windows Version 3.0
;; SEQ ID NO 483
;; LENGTH: 605
;; TYPE: DNA
;; ORGANISM: Homo sapiens
;; FEATURE:
;; NAME/KEY: misc_feature
;; LOCATION: (1)...(605)
;; OTHER INFORMATION: n = A,T,C or G
US-09-871-161-483
```

```
Query Match      18.4%; Score 187.2; DB 11; Length 605;
Best Local Similarity 65.5%; Pred. No. 5.6e-50;
Matches 348; Conservative 0; Mismatches 158; Indels 25; Gaps 5;
```

```
OY 485 GTACATGAGTAATTCACGTGAATTAAGTGCAGGGTAAAGCTTTTAACGGTTAAT 544
    ||||| || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 597 GGACAAGAAANNAAGTGCNGGTAAACCTTTAANGTAAATTTTGTCAATACATA 538
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 545 TCTGTCAACAGCTAG-ATGACAATGGCCGATCTTATCAGTGTCTCTTGACCCCT 603
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 537 GATTAANNAANNAAGTTGATTAACAGNNAATGTTTCTTANGCTTTTNTTNTCT 478
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 604 TCCCCCTGCTCCCTCCAGATGGGCGTTGAGTCAATTAATTAAGCGCATCCCA 663
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 477 AACACCTGCCATGCTCCCAATGGGCGATTTATTTATTTTAACNGTGTCTGT 418
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 664 CAGTGTCACTAGCAAGTCTTTCTTTAGACCCCTTTTAACGACAAATATGCT 723
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 417 TAGTCGTAACCTAGTGTAGTCTTTCTTTATAGAACCCCTTGCTGACGAATATGCT 358
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 724 GACCTGACTAATAGATTTCTGATATGATTCGAGATTTTGTGATAGTGA 783
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 357 -CCTGTATTAATAAATCTTTCTGATTAATGATTAGAGGTTTCTCGATTAGTAA 299
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 784 AGTCGCTTCTGTTTTCACCTTCTTACCTGAGTACTAGTCTGCCCTGCTTCTGA 843
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 298 AGTCCTTCCATGTAC-----TTTATTCAGAGCTAATAAGTCTTCTTAGTTTCTA 244
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 844 GTAACGGGTGTAGAAATACAGTGTGCGGCTTTACAGTTTAACTAATTTAGATA- 901
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 243 GTAACGTAGGTGTAAATAATCATGTGTGAGCTTTATAGTTTAAATTTTATAGTAAT 184
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 902 -----TCTGAACAATCAGTCTTTCGCCAGGTACCAACATGTCATGTG 946
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 183 TCTTAACATAGAACCTTCTTAACATCTGTCTTGCCAGATTCACGACATGTCACCTG 124
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 947 ATTGATGCGCGCCCTCTAGACCTACCCAGCGGACAGATGTTCCGGTA 997
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 123 ACCAATAGCTG-ACCCTCTTACCTGCGCCAGCGGACACACGCGCTCTGTGA 74
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

```
RESULT 14
US-10-066-543-3022/C
;; Sequence 3022, Application US/10066543
;; Publication No. US20030087818a1
;; GENERAL INFORMATION:
;; APPLICANT: Jjiang, Yugu
;; APPLICANT: Pyle, Ruth A.
;; APPLICANT: Xu, Jiaqun
;; APPLICANT: Indrias, Carol Joseph
;; APPLICANT: Lodes, Michael J.
;; APPLICANT: Secrist, Heather
```

```
;; APPLICANT: Carter, Darick
;; APPLICANT: Fanger, Gary R.
;; APPLICANT: Smith, Carole L.
;; APPLICANT: Durham, Margarita
;; APPLICANT: Stolk, John A.
;; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
;; TITLE OF INVENTION: AND DIAGNOSIS OF COLON CANCER
;; FILE REFERENCE: 210121.563
;; CURRENT APPLICATION NUMBER: US/10/066.543
;; CURRENT FILING DATE: 2002-01-31
;; NUMBER OF SEQ ID NOS: 3417
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 3022
;; LENGTH: 477
;; TYPE: DNA
;; ORGANISM: Homo sapiens
US-10-066-543-3022
```

```
Query Match      18.2%; Score 185.4; DB 14; Length 477;
Best Local Similarity 73.2%; Pred. No. 1.8e-49;
Matches 301; Conservative 0; Mismatches 86; Indels 24; Gaps 4;
```

```
OY 604 TCCCCCTGCTCCCTCCAGATGGGCGTTGAGTCAATTAATTAAGCGCATCCCTCA 663
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 477 TACACCTGCCATGCTCCCAATGGGCGATTTATTTATTTTAACGTGTGTCTGT 418
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 664 CAGTGTCACTAGCAAGTCTTTCTTTAGACCCCTTTTAACGACAAATATGCT 723
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 417 TAGTCGTAACCTAGTGTAGTCTTTCTTTATAGAACCCCTTGCTGACGAATATGCT 358
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 724 GACCTGACTAATAGATTTCTGATTAATGATTCGAGATTTTGTGATAGTGA 783
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 357 -CCTGTATTAATAAATCTTTCTGATTAATGATTAGAGGTTTGTGATAGTAA 299
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 784 AGTCGCTTCTGTTTTCACCTTCTTACCTGAGTACTAGTCTTCCCTGCTTCTGA 843
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 298 AGTCCTTCCATGTAC-----TTTATTCAGAGCTAATAAGTCTTCTTAGTTTCTA 244
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 844 GTAACGGGTGTAGAAATACAGTGTGCGGCTTTACAGTTTAACTAATTTAGATA- 901
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 243 GTAACGTAGGTGTAAATAATCATGTGTGAGCTTTATAGTTTAAATTTTATAGTAAT 184
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 902 -----TCTGAACAATCAGTCTTTCGCCAGATCAACCACTGTCATGTG 946
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 183 TCTTAACATAGAACCTTCTTAACATCTGTGTGCGGAGATTCACGACATGTCACCTG 124
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 947 ATTGATGCGCGCCCTCTAGACCTACCCAGCGGACAGATGTTCCGGTA 997
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 123 ACCAATAGCTG-ACCCTCTTACCTGCGCCAGCGGACACACGCGCTCTGTGA 74
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

```
RESULT 15
US-09-960-352-7612
;; Sequence 7612, Application US/09960352
;; Patent No. US20020137139A1
;; GENERAL INFORMATION:
;; APPLICANT: Warren, Wesley C.
;; APPLICANT: Tao, Nengping
;; APPLICANT: Byatt, John C.
;; APPLICANT: Mathialagan, Nagappan
;; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
;; FILE REFERENCE: 16511.006/37-21(10298)C
;; CURRENT APPLICATION NUMBER: US/09/960.352
;; CURRENT FILING DATE: 2001-09-24
;; NUMBER OF SEQ ID NOS: 15112
;; SEQ ID NO 7612
;; LENGTH: 432
;; TYPE: DNA
;; ORGANISM: Bos taurus
;; OTHER INFORMATION: Clone ID: 33-LIB188-020-01-EL-A2
US-09-960-352-7612
```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 25, 2003, 21:56:10 ; Search time 1283.94 Seconds

(without alignments)
11088.207 Million cell updates/sec

Title: US-09-717-321A-17

Perfect score: 348
Sequence: 1 tgaacatcactgtcttgc.....cttgaacttgaaaaaa 348

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :
1: GenBank:
2: gb_da:
3: gb_in:
4: gb_lm:
5: gb_ov:
6: gb_pat:
7: gb_ph:
8: gb_pl:
9: gb_pr:
10: gb_ro:
11: gb_sts:
12: gb_sy:
13: gb_un:
14: gb_vl:
15: em_da:
16: em_fun:
17: em_hum:
18: em_in:
19: em_mu:
20: em_om:
21: em_or:
22: em_ov:
23: em_pat:
24: em_ph:
25: em_pl:
26: em_ro:
27: em_sts:
28: em_un:
29: em_vl:
30: em_hgtg_hum:
31: em_hgtg_inv:
32: em_hgtg_other:
33: em_hgtg_mus:
34: em_hgtg_pln:
35: em_hgtg_rnd:
36: em_hgtg_mam:
37: em_hgtg_vrt:
38: em_sy:
39: em_hgtgo_hum:
40: em_hgtgo_mus:
41: em_hgtgo_other:

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	348	100.0	348	6	AX163753	AX163753 Sequence
2	348	100.0	1268	6	AX163738	AX163738 Sequence
3	320	92.0	217700	2	AC106124	AC106124 Rattus no
4	230.2	66.1	2281	10	BC051053	BC051053 Mus muscu
5	228.2	65.6	2319	10	BC003828	BC003828 Mus muscu
6	221	63.5	269081	2	AC068493	AC068493 Mus muscu
7	216.8	62.3	192498	2	AC105979	AC105979 Mus muscu
8	216.8	62.3	230015	2	AC132602	AC132602 Mus muscu
9	207.4	58.6	2302	9	BC050687	BC050687 Homo sapi
10	203.6	58.5	28567	9	HSN132695	AJ132695 Homo sapi
11	203.6	58.5	212827	9	AC009412	AC009412 Homo sapi
12	192.4	55.3	2315	9	AK054993	AK054993 Homo sapi
13	187.8	54.0	352	6	AX163754	AX163754 Sequence
14	187.4	53.9	1232	6	AXA08985	AXA08985 Sequence
15	187.4	53.9	1232	9	HDMPO2ST9	D25274 Homo sapien
16	176.4	50.7	87616	2	AC139405	AC139405 Homo sapi
17	176.4	50.7	137625	9	AC104663	AC104663 Homo sapi
18	175.6	50.5	240973	2	AC123247	AC123247 Rattus no
19	175.6	50.5	241048	2	AC129824	AC129824 Rattus no
20	174.8	50.2	5544	9	AF542527	AF542527 Homo sapi
21	170.6	49.0	262	11	G31709	G31709 swiss2233 Er
22	154	44.3	455	11	G26995	G26995 human STS S
23	123	35.3	123	6	AX163737	AX163737 Sequence
24	121.4	34.9	228121	2	AC133022	AC133022 Rattus no
25	121.4	34.9	239768	2	AC112582	AC112582 Rattus no
26	121.4	34.9	245668	2	AC130746	AC130746 Rattus no
27	116.4	33.4	110816	9	AC002404	AC002404 Human Chr
28	113	32.5	1017	6	AX163751	AX163751 Sequence
29	110.6	31.8	101584	9	CNS01DS5	AL121655 BAC sequ
30	110.6	31.8	155943	2	AC012364	AC012364 Homo sapi
31	84.6	24.3	174316	2	AC022647	AC022647 Homo sapi
32	84.6	24.3	217249	9	AC009902	AC009902 Homo sapi
33	81.2	23.3	118648	9	AL138742	AL138742 Human DNA
34	81.2	23.3	131095	2	AL590071	AL590071 Homo sapi
35	79.2	22.8	170839	2	AC133467	AC133467 Mus muscu
36	75.2	21.6	237985	2	AC125754	AC125754 Rattus no
37	75.2	21.6	247478	2	AC097964	AC097964 Rattus no
38	72	20.7	156879	10	AL626786	AL626786 Mouse DNA
39	67	19.3	340701	2	AC120633	AC120633 Rattus no
40	66.8	19.2	62856	2	AC100110	AC100110 Mus muscu
41	66.8	19.2	178482	2	AC101810	AC101810 Mus muscu
42	60.8	17.5	64	6	AX522492	AX522492 Sequence
43	53.4	15.3	100697	5	AL672072	AL672072 zebrafish
44	52.4	15.1	162	6	AX397334	AX397334 Sequence
45	52.4	15.1	1022	6	AX163752	AX163752 Sequence

ALIGNMENTS

RESULT 1
LOCUS AX163753 348 bp
DEFINITION Sequence 17 from Patent WO0138579.
ACCESSION AX163753
VERSION AX163753.1 GI:14544859
KEYWORDS
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1
Gould-Rothberg, B.E., DiPippo, V.A., Ramsehl, T.M. and Gerweil, R.W.
Method of Identifying Toxic Agents using nsaid-induced differential

gene expression in liver
 Patient: WO 0138579-A 17 31-MAY-2001;
 Curagen Corporation (US)
 Location/Qualifiers
 FEATURES
 source
 1.348
 /organism="Rattus norvegicus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10116"

BASE COUNT 103 a 76 c 67 g 102 t

ORIGIN

Query Match 100.0%; Score 348; DB 6; Length 348;
 Best Local Similarity 100.0%; Pred. No. 1.8e-78;
 Matches 348; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGAACATCATCTGCTTGGCAGAGTACCAACACTGTCATGTGATGGCCGCCCTCT 60
 |||||||
 DB 1 TGAACATCATCTGCTTGGCAGAGTACCAACACTGTCATGTGATGGCCGCCCTCT 60
 |||||||

QY 61 AGACCTCACCACGCGGACACATGCTTCGGGTACCTTTGGGTCTGTGAGTTCTGTCAAG 120
 |||||||
 DB 61 AGACCTCACCACGCGGACACATGCTTCGGGTACCTTTGGGTCTGTGAGTTCTGTCAAG 120
 |||||||

QY 121 CGCTAGTGTCTAACGCCGCTTCTGTACAACCTACTGCTGAGCAAGAACAGTGTGGGCC 180
 |||||||
 DB 121 CGCTAGTGTCTAACGCCGCTTCTGTACAACCTACTGCTGAGCAAGAACAGTGTGGGCC 180
 |||||||

QY 181 TTTCGACCACTAGAACAACTTTTTCATTTGACAGTTGCGAATTTGGAGTGTCTTTTA 240
 |||||||
 DB 181 TTTCGACCACTAGAACAACTTTTTCATTTGACAGTTGCGAATTTGGAGTGTCTTTTA 240
 |||||||

QY 241 CATTGATCTTTTCTTAATGACATGATGATGATGATGATGATGATGATGATGATGATGAT 300
 |||||||
 DB 241 CATTGATCTTTTCTTAATGACATGATGATGATGATGATGATGATGATGATGATGATGAT 300
 |||||||

QY 301 TGAATCATATAAAAAAAAAAATGTCTTTGGAATTTGAAAAAAA 348
 |||||||
 DB 301 TGAATCATATAAAAAAAAAAATGTCTTTGGAATTTGAAAAAAA 348
 |||||||

RESULT 2
 LOCUS AX163738 1266 bp DNA linear PAT 22-JUN-2001
 DEFINITION Sequence 2 from Patent WO0138579.
 AX163738
 VERSION AX163738.1 GI:14544844
 KEYWORDS
 SOURCE Rattus norvegicus (Norway rat)
 ORGANISM Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.

REFERENCE
 1 Gould-Rothberg, B.E., Dipippo, V.A., Ramsehl, T.M. and Gerwehn, R.W.
 Method of identifying toxic agents using nsaid-induced differential
 gene expression in liver
 Patent: WO 0138579-A 2 31-MAY-2001;
 Curagen Corporation (US)

FEATURES
 source
 1.1266
 /organism="Rattus norvegicus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10116"

BASE COUNT 385 a 258 c 285 g 338 t

ORIGIN

Query Match 100.0%; Score 348; DB 6; Length 1266;
 Best Local Similarity 100.0%; Pred. No. 2e-78;
 Matches 348; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGAACATCATCTGCTTGGCAGAGTACCAACACTGTCATGTGATGGCCGCCCTCT 60
 |||||||
 DB 357 TGAACATCATCTGCTTGGCAGAGTACCAACACTGTCATGTGATGGCCGCCCTCT 298
 |||||||

QY 61 AGACCTCACCACGCGGACACATGCTTCGGGTACCTTTGGGTCTGTGAGTTCTGTCAAG 120
 |||||||
 DB 297 AGACCTCACCACGCGGACACATGCTTCGGGTACCTTTGGGTCTGTGAGTTCTGTCAAG 238
 |||||||

QY 121 CGCTAGTGTCTAACGCCGCTTCTGTACAACCTACTGCTGAGCAAGAACAGTGTGGGCC 180
 |||||||
 DB 237 CGCTAGTGTCTAACGCCGCTTCTGTACAACCTACTGCTGAGCAAGAACAGTGTGGGCC 178
 |||||||

QY 181 TTTCGACCACTAGAACAACTTTTTCATTTGACAGTTGCGAATTTGGAGTGTCTTTTA 240
 |||||||
 DB 177 TTTCGACCACTAGAACAACTTTTTCATTTGACAGTTGCGAATTTGGAGTGTCTTTTA 118
 |||||||

QY 241 CATTGATCTTTTCTTAATGACATGATGATGATGATGATGATGATGATGATGATGATGAT 300
 |||||||
 DB 117 CATTGATCTTTTCTTAATGACATGATGATGATGATGATGATGATGATGATGATGATGAT 58
 |||||||

QY 301 TGAATCATATAAAAAAAAAAATGTCTTTGGAATTTGAAAAAAA 348
 |||||||
 DB 57 TGAATCATATAAAAAAAAAAATGTCTTTGGAATTTGAAAAAAA 10
 |||||||

RESULT 3
 LOCUS AC106124 217700 bp DNA linear HTG 10-MAY-2003
 DEFINITION Rattus norvegicus clone CH230-119E11, *** SEQUENCING IN PROGRESS
 AC106124
 VERSION AC106124.6 GI:30521557
 KEYWORDS HTG: HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
 SOURCE Rattus norvegicus (Norway rat)
 ORGANISM Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.

REFERENCE
 1 (bases 1 to 217700)
 Muzny, D., Marle, J., Metzger, M., Lee, A., Abramson, S., Adams, C., Alder, J.,
 Allen, C., Allen, H., Ashbrook, S., Amin, A., Anguiano, D.,
 Anyalebech, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,
 Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benhmed, F.,
 Biswal, K., Blair, J., Blankenburg, R., Blyth, P., Brown, M.,
 Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,
 Cardenas, V., Carter, K., Cavazos, I., Caesar, H., Center, A.,
 Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,
 Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Grege, A., D'Souza, L.,
 Davila, M.L., Davis, C., Davy-Carroll, L., De Anda, C., Dedrich, D.,
 Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Dlyva, K.,
 Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,
 Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G.,
 Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,
 Fraser, C.M., Gablitz, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,
 Gebregeorgis, E., Geer, K., Gill, R., Greedy, M., Guerra, W., Guevara, W.,
 Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K.,
 Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,
 Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogue, M.,
 Hollins, B., Howells, S., Hulik, S., Hume, J., Idelbird, D., Jackson, A.,
 Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolyet, A.,
 Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, R., Kovar, C.,
 Kowis, C., Kraft, C.L., Ledow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,
 Liu, J., Liu, M., Liu, Y., London, P., Longacre, S., Lopez, J.,
 Lorensuwa, L., Louisedge, H., Lozano, R.J., Lu, X., Ma, J.,
 Maheshwari, M., Mahlodierne, M., Mahmoud, M., Malloy, K., Mangum, A.,
 Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,
 Mahoney, S., McLeod, M.P., McNeill, T.Z., Meenan, E.,
 Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,
 Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L.,
 Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,
 Nwankweme, O., Okunolu, G., Olarinmoye, A., Pal, S., Parks, K.,
 Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C.,
 Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L., -L.,
 Puzo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R.,
 Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,
 Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J.,

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
Sanders,W., Savary,G., Scherer,S., Scott,G., Shatman,S., Shen,H., Shetty,J., Shivatsbeyan,A., Slasson,I., Sitter,C.D., Smajs,D., Sneed,A., Sodergren,E., Song,X.-Z., Sorreller,R., Sosa,J., Steimle,M., Strong,R., Sutton,A., Swalek,A., Tabor,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K., Valas,R., Vera,V., Villasaena,D., Waldron,L., Walker,B., Wang,J., Wang,O., Wang,S., Warren,J., Warren,R., Wei,X., White,F., Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V., Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O., Melnicko,G. and Gibbs,R.A.
Direct Submission
Unpublished
2 (bases 1 to 217700)
Worley,K.C.
Direct Submission
Submitted (12-JAN-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 217700)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (10-MAY-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On May 10, 2003 this sequence version replaced gi:25094619. The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/ratl/). Each contig described in the feature table below represents a scaffold in the Atlas assembly ('a' contig-scaffold). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.
----- Genome Center Center: Baylor College of Medicine Center code: BCM Web site: http://www.hgsc.bcm.tmc.edu/ Contact: hgsc-help@bcm.tmc.edu Project Information Center project name: CHIL Center clone name: CH230-119E11 ----- Summary Statistics Assembly program: Atlas 3.0; Consensus quality: 185725 bases at least Q40 Consensus quality: 190140 bases at least Q30 Consensus quality: 192842 bases at least Q20 Estimated insert size: 194755; sum-of-contigs estimation Quality coverage: 5x in Q20 bases; sum-of-contigs estimation ----- * NOTE: Estimated insert size may differ from sequence length * (see http://www.hgsc.bcm.tmc.edu/docs/genbank_drift_data.html) * NOTE: This is a 'working draft' sequence. It currently consists of 11 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. * This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved. * * 1 109797: contig of 109797 bp in length * 109798 109897: gap of unknown length * 109898 166934: contig of 57037 bp in length * 166935 167034: gap of unknown length * 167035 176384: contig of 9350 bp in length * 176385 176484: gap of unknown length * 176485 206327: contig of 28643 bp in length * 206328 206427: gap of unknown length

	*	206428	207461:	contig of 1034 bp in length	
	*	207462	207561:	gap of unknown length	
	*	207562	208607:	contig of 1046 bp in length	
	*	208608	208707:	gap of unknown length	
	*	208708	209908:	contig of 1201 bp in length	
	*	209909	210008:	gap of unknown length	
	*	210009	211583:	contig of 1575 bp in length	
	*	211584	211683:	gap of unknown length	
	*	211684	213649:	contig of 1966 bp in length	
	*	213650	213749:	gap of unknown length	
	*	213750	215115:	contig of 1366 bp in length	
	*	215116	215215:	gap of unknown length	
	*	215216	217700:	contig of 2485 bp in length.	
FEATURES					
source					
			1..217700		
			/organism="Rattus norvegicus"		
			/mol_type="genomic DNA"		
			/db_xref="taxon:10116"		
			/clone="CH230-119E11"		
BASE COUNT		51384 a	45640 c	45954 g	51476 t 23246 others
ORIGIN					
Query Match			92.0%:	Score 320;	DB 2; Length 217700;
Best Local Similarity			95.6%:	Pred. No. 3.6e-71;	
Matches 329; Conservative			0; Mismatches 15;	Indels 0; Gaps 0;	
OY	1	TGAACATCATCTGTCCTTGGCCAGAGTACCACAACACTGTGCATGTGATGATGCCCGCCCTCT	60		
Dd	64351	TGAACATCATCTGTCCTTGGCCAGAGTACCACAACACTGTGCATGTGATGATGCCCGCCCTCT	64410		
OY	61	AGACCTTAGCCAGCGGAGACATGCTTCGGGTAACCTTTGGGTCGTGAGGTTCTGTCAAG	120		
Dd	64411	AGACCTTAGCCAGCGGAGACATGCTTCGGGTAACCTTTGGGTCGTGAGGTTCTGTCAAG	64470		
OY	121	CGTAGTGCTAAGCGCCTTCTGTACAACTCACTGACGCGCAAGACAGAGTGTGGGCGC	180		
Dd	64471	CGTAGTGCTAAGCGCCTTCTGTACAACTCACTGACGCGCAAGACAGAGTGTGGGCGC	64530		
OY	181	TTTTGCACCTAGAACAACTTTTTCATTTGACAGTTCAGAAATTGTGGAGTGTTTTAA	240		
Dd	64531	TTTTGCACCTAGAACAACTTTTTCATTTGACAGTTCAGAAATTGTGGAGTGTTTTAA	64590		
OY	241	CATTGATCTTTTGGCTAATGCAATTAGCAGTAGTATTTTGCATGTATTAATAATCCT	300		
Dd	64591	CATTGATCTTTTGGCTAATGCAATTAGCAGTAGTATTTTGCATGTATTAATAATCCT	64650		
OY	301	TGATTCATTAATAAAAAAAAAAAAAATGCTTTTGGAACTTGAANA 344			
Dd	64651	TGATTCATTAATAACTGTGTCATGCTTGTGCTTTGGAACTTGAANA 64694			
RESULT 4					
BC051053					
LOCUS		BC051053	2281 bp	mRNA	linear ROD 14-APR-2003
DEFINITION		Mus musculus, clone MGC:58966 IMAGE:5038182, mRNA, complete cds.			
ACCESSION		BC051053			
VERSION		BC051053.1	GI:29835221		
KEYWORDS		MGC.			
SOURCE		Mus musculus (house mouse)			
ORGANISM		Mus musculus			
		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
		Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
		1 (bases 1 to 2281)			
REFERENCE					
AUTHORS		Strausberg,R.			
TITLE		Direct Submission			
JOURNAL		Submitted (11-APR-2003) National Institutes of Health, Mammalian			
		Gene Collection (MGC), Cancer Genomics Office, National Cancer			
		Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,			
		USA			
REMARK		NIH-MGC Project URL: http://mgc.ncl.nih.gov			
COMMENT		Contact: MGC help desk			
		Email: gcgabs-remail.nih.gov			
		Tissue Procurement: Gilbert Smith, Ph.D.			

gene

CDS

ORIGIN

BASE COUNT 603 a 573 c 503 g 640 t

Query Match 65.6%; Score 228.2; DB 10; Length 2319;
Best Local Similarity 83.7%; Pred. No. 8.5e-48;
Matches 292; Conservative 0; Mismatches 33; Indels 24; Gaps 2;

OY 21 AGACTACCAACATGTCATGTGAATTGGTAAGCCGCCCTTGTAGACTCACCACGCAGCA 80
Db 1964 ACACTACCAACAACATGTCATGTGACTAATGCG---CCCTTGACCTCCGCCACGTGACCA 2020
OY 81 CATGCTTCGGGTA-----CCTTTGGGCTGTGAGGTCTGTCAA 119
Db 2021 GAGCCTTCCTGTACATNGGCTCTGCCTTAGAGATGTTCTTGGSGTCTGTGAGTTCTGTCAA 2088
OY 120 GCCTAGTGTCTAACGCCGTGCTGTACAACTTACTACTGTGGCAAGACAGTGTGGGC 179
Db 2081 CTGCTAGTGTCTAACGCCGTGCTGTACAACTTACTACTGTGGCAAGATGCAGTGTGGGC 2140
OY 180 CTTTTGACCATATAACAACACTTTTTCATATGACAGTGCAGAATTGTGGAGTGTTTY 239
Db 2141 CTCCTAACCATATAACAACACTTTTTCATATGACAGTGCAGAATTGTGGAGTGTTTY 2200
OY 240 ACATTGATCTTTTGTCTATGACAGTGTAGCAGTATGTTTGATGATGACTTAATAATCC 299
Db 2201 ACAAATGATCTTTTGTCTATGACAGTGTAGCAGTATGTTTGATGATGATTAATAAATCC 2260
OY 300 TTGATCATATAAAAAAAAAAAAAATGCTTTGGAACTGAAAANANA 348
Db 2261 TTTGATCATATAAAAAAAAAAAAAATGCTTTGGAACTGAAAANANA 2309

RESULT 6
LOCUS 269081 bp DNA linear HTG 15-MAY-2002Z
ACO68493 Mus musculus clone RP23-7K1 strain C57BL6/J, WORKING DRAFT
DEFINITION
SEQUENCE, 26 unordered pieces.
ACCESSION
VERSION ACO68493.10 GI:15148081
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE Mus musculus (house mouse)
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
AUTHORS Montgomey,K.T., Grillis,G., Han,J., Lee,E., Long,J., Pomerantz,R.,
Ioshihes,I.P., Shlm,C., Decker,J., Thomas,E., Peters,A.,
Gordon,M., Goltz,J.S. and Kucheriapatti,R.
TITLE High Throughput Mouse Sequencing
JOURNAL
REFERENCE Unpublished
AUTHORS 2 (bases 1 to 269081)
Montgomey,K.T., Grillis,G., Han,J., Lee,E., Long,J., Pomerantz,R.,
Ioshihes,I.P., Shlm,C., Decker,J., Thomas,E., Peters,A.,
Gordon,M., Goltz,J.S. and Kucheriapatti,R.
TITLE Direct Submission
JOURNAL
SUBMITTED (03-MAY-2000) Department of Molecular Genetics, Albert

COMMENT

* NOTE: This is a 'working draft' sequence. It currently consists of 26 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.	
1	51730: contig of 51730 bp in length
51731	51750: gap of unknown length
51751	82423: contig of 30673 bp in length
82424	82443: gap of unknown length
82444	118158: contig of 35715 bp in length
118159	118178: gap of unknown length
118179	139441: contig of 21263 bp in length
139442	139461: gap of unknown length
139462	160638: contig of 21177 bp in length
160639	160658: gap of unknown length
160659	177611: contig of 16953 bp in length
177612	177631: gap of unknown length
177632	190991: contig of 13360 bp in length
190992	191011: gap of unknown length
191012	200128: contig of 9117 bp in length
200129	200148: gap of unknown length
200149	211882: contig of 11734 bp in length
211883	211902: gap of unknown length
211903	220773: contig of 8871 bp in length
220774	220793: gap of unknown length
220794	229643: contig of 8850 bp in length
229644	229663: gap of unknown length
229664	238339: contig of 8676 bp in length
238340	238359: gap of unknown length
238360	244848: contig of 6489 bp in length
244849	244868: gap of unknown length
244869	249725: contig of 4857 bp in length
249726	249745: gap of unknown length
249746	252523: contig of 2778 bp in length
252524	252543: gap of unknown length
252544	257520: contig of 4977 bp in length
257521	257540: gap of unknown length
257541	260396: contig of 2856 bp in length
260397	260416: gap of unknown length
260417	262709: contig of 2293 bp in length
262710	262729: gap of unknown length
262730	263934: contig of 1205 bp in length
263935	263954: gap of unknown length
263955	264196: contig of 242 bp in length
264197	264216: gap of unknown length
264217	264629: contig of 413 bp in length
264630	264649: gap of unknown length
264650	265608: contig of 959 bp in length
265609	265628: gap of unknown length
265629	265935: contig of 1307 bp in length

Cook, A., Cooke, P., DeRellano, K., Dewar, K., Diaz, J. S., Dodge, S.,
 Fero, S., Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J.,
 Gardy, S., Gord, S., Graham, L., Grand-Pierre, N., Hafez, N.,
 Haggos, B., Horton, L., Hulme, L., Illey, I., Johnson, R., Jones, C.,
 Kama, A., Karatas, A., Kells, C., Landers, T., Levine, R.,
 Lindblad-TO, K., Liu, G., Maclean, C., MacDonald, P., Major, J.,
 Matthews, C., McCarthy, M., Meldrum, J., Menues, L., Mihova, T.,
 Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C.,
 Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J.,
 Peterson, K., Phunkhang, P., Plerre, N., Raymond, C., Retta, R.,
 Rise, C., Rogov, P., Roman, J., Roy, A., Schaner, S., Schupbach, R.,
 Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N.,
 Stojanovic, N., Talamas, J., Testaye, S., Theodore, J., Toppan, K.,
 Travers, M., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X.,
 Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
 Direct Submission
 Submitted (21-OCT-2002) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Oct 21, 2002 this sequence version replaced g1:20149424.

```

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: MIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
-----
Project Information
Center project name: L20199
Center clone name: 301_E_12
-----
Summary Statistics
Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 190070 bases at least Q40
Consensus quality: 191243 bases at least Q30
Consensus quality: 191514 bases at least Q20
Insert size: 163000; agarose-fp
Insert size: 191698; sum-of-contigs
Quality coverage: 11.6 in Q20 bases; agarose-fp
Quality coverage: 9.9 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 9 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* been provided by the submittor.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
*
1
* 1062 1061: contig of 1061 bp in length
* 1162 1161: gap of 100 bp
* 5715 5714: contig of 4553 bp in length
* 5815 5814: gap of 100 bp
* 11423 11422: contig of 5608 bp in length
* 11523 11522: gap of 100 bp
* 17601 17600: contig of 6078 bp in length
* 17701 17700: gap of 100 bp
* 30158 30157: contig of 12457 bp in length
* 30258 30257: gap of 100 bp
* 38788 38787: contig of 8530 bp in length
* 38888 38887: gap of 100 bp
* 99444 99443: contig of 60557 bp in length
* 99545 99544: gap of 100 bp
* 143263 143262: contig of 43719 bp in length
* 143264 143263: gap of 100 bp
* 143364 143363: contig of 49135 bp in length.
*
location/Qualifiers
1. 192498
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
FEATURES
source

```

	misc_feature	/clone="RP24-301E12"	
		/clone_1lb="RPCT-24 Male Mouse BAC"	
	misc_feature	1..1061	
	misc_feature	/note="assembly_fragment"	
		1162..5714	
	misc_feature	/note="assembly_fragment"	
		5815..11422	
	misc_feature	/note="assembly_fragment"	
		11523..17600	
	misc_feature	/note="assembly_fragment"	
		17701..30157	
	misc_feature	/note="assembly_fragment"	
		30258..38787	
	misc_feature	/note="assembly_fragment"	
		38888..99444	
	misc_feature	/note="assembly_fragment"	
		99545..143263	
	misc_feature	/note="assembly_fragment"	
		143364..192498	
	misc_feature	/note="assembly_fragment"	
		vector_end:"T7"	
		vector_side:"light"	
BASE COUNT	53596 a 42471 c 43252 g 52379 t	800 others	
ORIGIN			
Query Match	62.3%; Score 216.8; DB 2;	Length 192498;	
Best Local Similarity	86.1%; Pred. No. 8.6e-45;		
Matches 272; Conservative	0; Mismatches 22; Indels 22; Gaps 2;		
OY	29 AACACGTGATGATTGATGGCGCCGCCCTTGAGACTCACCCAGCGGACACATGCTTC	88	
Db	90554 AACCCTGTCAATGACTAATGCTG-CCTCTTAGACCTCCTACAGAGACAGCTTC	90496	
OY	89 CGGTA-----CCTTTGGCTGTGAGGTTCTGTCAAGCGTAGT	127	
Db	90495 CTGTGCCTGGCTGCGCCAGAGATGTTCTTGGGGTGTGAGGTTCTGTCAACTGTAAT	90433	
OY	128 GCTAACGCCCTTCTGTACAACTACTACTGCGCAAGAACAACAGTGTGGCCCTTTCGAC	187	
Db	90435 GCTAATCTGTCCGTGCAACCTCACTCACTGCGCAGAAATACAGTGTGGCCCTTTCAC	90376	
OY	188 CACTAGAACAACTTTTTCATATGACAGTGCAGAAATTTGGAGTGTTTTACATGTAT	247	
Db	90375 CACTAGAACAACTTTTTCATATGACAGTGCAGAAATTTGGAGTGTTTTACATGTAT	90318	
OY	248 CTTTTCGTAATGACAGTATGATTTTTCATGTATGATGACTTAATAATCCTGAATCA	307	
Db	90315 CTTTTCGTAATGACAGTATGATTTTTCATGTATGATGACTTAATAATCCTGAATCA	90256	
OY	308 TAAAAAAAAGAAAA 323		
Db	90255 TTTAAAAAAAAGAAA 90240		
RESULT 8			
AC132602/c			
LOCUS	AC132602	230015 bp	DNA linear HTG 01-SEP-2002
DEFINITION	Mus musculus chromosome UNK clone RP23-477Cl3, WORKING DRAFT		
ACCESSION	AC132602		
VERSION	AC132602.1	GI:22597688	
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.		
SOURCE	Mus musculus (house mouse)		
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
REFERENCE	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
AUTHORS	McPherson,J.D. and Waterston,R.H.		
JOURNAL	The sequence of Mus musculus clone		
REFERENCE	Unpublished		
AUTHORS	McPherson,J.D. to 230015)		
TITLE	Direct Submission		

JOURNAL Submitted (01-SEP-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA

COMMENT

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: <http://genome.wustl.edu/gsc/index.shtml>
Contact: submissions@wustl.wustl.edu
----- Project Information -----
Center project name: M_BA0477C13
----- Summary Statistics -----
Sequencing vector: MJ3: 0%
Chemistry: Dye-Primer ET; 0% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 225083 bases at least Q40
Consensus quality: 225730 bases at least Q30
Consensus quality: 226267 bases at least Q20
Insert size: 20200; agarose-1p
Insert size: 228405; sum-of-ctrls
Quality coverage: 13.11 in Q20 bases; agarose-1p
Quality coverage: 9.64 in Q20 bases; sum-of-ctrls

* NOTE: This is a 'working draft' sequence. It currently
* consists of 10 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```

1      710: contig of 710 bp in length
*      711      810: gap of unknown length
*      811      18316: contig of 17506 bp in length
*      18317      18416: gap of unknown length
*      18417      57380: contig of 38964 bp in length
*      57381      57480: gap of unknown length
*      57481      118173: contig of 60693 bp in length
*      118174      118273: gap of unknown length
*      118274      201518: contig of 83245 bp in length
*      201519      201618: gap of unknown length
*      201619      202801: contig of 1183 bp in length
*      202802      202901: gap of unknown length
*      202902      204714: contig of 1813 bp in length
*      204715      204814: gap of unknown length
*      204815      208091: contig of 3277 bp in length
*      208092      208191: gap of unknown length
*      208192      213082: contig of 4891 bp in length
*      213083      213182: gap of unknown length
*      213183      230015: contig of 16833 bp in length.

```

FEATURES

source

```

1. .230015
  /organism="Mus musculus"
  /mol_type="genomic DNA"
  /db_xref="taxon:10090"
  /chromosome="XUK"
  /clone="RP23-477C13"
  1. .710
    /note="assembly_name:Contig1"
    811. .18316
      /note="assembly_name:Contig10"
      18417. .57380
        /note="assembly_name:Contig11"
        57481. .118173
          /note="assembly_name:Contig12"
          118274. .201518
            /note="assembly_name:Contig13"
            201619. .202801
              /note="assembly_name:Contig3"
              202902. .204714
                /note="assembly_name:Contig6"

```

```

misc_feature      204815..208091
                  /note="assembly_name:Contig7"
misc_feature      208192..213082
                  /note="assembly_name:Contig8"
misc_feature      213183..230015
                  /note="assembly_name:Contig9"
BASE COUNT      63025 a 50551 c 50278 g 65213 t      948 others
ORIGIN

```

Query Match 62.3%; Score 216.8; DB 2; Length 230015;
Best local Similarity 86.1%; Pred. No. 8, 6e-45;
Matches 272; Conservative 0; Mismatches 22; Indels 22; Gaps 2;

```

QY      29  AACACGTGTCATGATGATGATGCCGCCCTTAGACCTGACCGCGGACACATGCTTC 88
          |||||
DB      109281 AACACGTGTCATGATGATGATGCTG-CCTCTGACCTGACCTGACGAGACAGACGCTTC 109223
          |||||
QY      89  CGGTA-----CCTTGGCTGTCGAGGTCTGTCACCGCTTACT 127
          |||||
DB      109222 CTGTGCGTGGCTGCTGCTGACGATGCTTCTGGGCTGTGAGGTCTGTCACCTGCTTACT 109163
          |||||
QY      128  GCTAACGCGTGTCTGTACACCTTACTGCTGCAAGAACACAGTGTGGCCTTTCGAC 187
          |||||
DB      109162 GCTAATGCTGTCTGTGTACCACTTACTGCTGCAAGAACATGAGTGTGGCCTTTCACAC 109103
          |||||
QY      188  CACTAAGACAACTTTTTCATTTGACAGTTCGACGATTTGAGATTTGAGTCTTTTACATTGAT 247
          |||||
DB      109102 CACTAAGACAACTTTTTCATTTGACAGTTCGACGATTTGAGATTTGAGTCTTTTACATTGAT 109043
          |||||
QY      248  CTTTTCCTAATGAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 307
          |||||
DB      109042 CTTTTCCTAATGAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 108983
          |||||
QY      308  TAAAAAATTTTTTTTTT 323
          |||||
DB      108982 TTTAAAAAATTTT 108967
          |||||

```

RESULT 9 2302 bp mRNA linear PRI 11-APR-2003
BC050687
LOCUS
DEFINITION
Homo sapiens, ras-related C3 botulinum toxin substrate 1 (rho
family, small GTP binding protein Rac1), clone MGC:60264
IMAGE:6149377, mRNA, complete cds.

ACCESSION BC050687.1 GI:29792301
VERSION BC050687.1
KEYWORDS MGC.

SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 2302)
Strausberg, R.

AUTHORS Direct Submission
TITLE Submitted (08-APR-2003) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA

REMARK NIH-MGC Project URL: <http://mgc.ncl.nih.gov>
Contact: MGC help desk
Email: cgabs-remail.nih.gov
Tissue Procurement: ATCC/DCPD/DMP

CDNA Library Preparation: Life Technologies, Inc.

DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

Center, Stanford University School of Medicine, Stanford, CA 94305

Web site: <http://www-ehg.stanford.edu>

Contact: (Dickson, Mark) medepaxl@stanford.edu

Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,

R. M.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>

1. .2302

FEATURES
source

```
repeat_region
repeat_region
```

1

11-11-11

וידועה - דכחמא

misc_feature

repeat_region

33

CCC

100

```

Intron      1262..13700
            /gene="rac1"
            /number=1
repeat_region /rpt_family="GC_r1ch"
            1390..1419
repeat_region /rpt_family="FLAM"
            1752..1793
repeat_region /rpt_family="AluJo"
            1794..2005
repeat_region /rpt_family="AluJo"
            2140..2171
repeat_region /rpt_family="Simple_repeat"
            2666..2972
repeat_region /rpt_family="Alusx"
            2980..3267
repeat_region /rpt_family="Alusx"
            3268..3289
repeat_region /rpt_family="(r)n"
            3802..4102
repeat_region /rpt_family="AluJb"
            4122..4386
repeat_region /rpt_family="Alusx"
            4538..4719
repeat_region /rpt_family="LTR33A"
            4754..5053
repeat_region /rpt_family="Alusq"
            5297..5594
repeat_region /rpt_family="Alusx"
            5655..5757
repeat_region /rpt_family="Alusq"
            5758..6062
repeat_region /rpt_family="Alusq"
            6063..6198
repeat_region /rpt_family="Alusq"
            6222..6401
repeat_region /rpt_family="Alusq/x"
            7162..7238
repeat_region /rpt_family="LMCS"
            7348..7367
repeat_region /rpt_family="Simple_repeat"
            7368..7656
repeat_region /rpt_family="Alusx"
            7880..8183
repeat_region /rpt_family="Alusx"
            8484..8778
repeat_region /rpt_family="Alusx"
            9071..9365
repeat_region /rpt_family="Alusx"
            10279..10583
repeat_region /rpt_family="Alusq"
            10609..10864
repeat_region /rpt_family="L2"
            10970..11276
repeat_region /rpt_family="AluJo"
            11276..11594
repeat_region /rpt_family="AluJb"
            11744..11825
repeat_region /rpt_family="L2"
            11764..11987
repeat_region /rpt_family="MIR"
            12989..13307
repeat_region /rpt_family="Alusx"
            13701..13772
exon         /gene="rac1"
            /number=2
            13773..18412
Intron       /gene="rac1"
            /number=2
            13954..14235
repeat_region /rpt_family="Alusq"
            14807..14831
repeat_region /rpt_family="Simple_repeat"
            14836..15128
repeat_region /rpt_family="Alusx"

```

```

repeat_region 15408..15715
            /rpt_family="Alusx"
repeat_region 15872..15896
            /rpt_family="AT_r1ch"
repeat_region 15902..16186
            /rpt_family="Alusx"
repeat_region 16187..16314
            /rpt_family="AluJo"
repeat_region 16413..16585
            /rpt_family="Simple_repeat"
repeat_region 16586..16679
            /rpt_family="FLAM_A"
repeat_region 16707..16810
            /rpt_family="LM4"
repeat_region 16811..17108
            /rpt_family="Alusq"
repeat_region 17109..17345
            /rpt_family="LM4"
repeat_region 17377..17686
            /rpt_family="Alusq"
repeat_region 17703..17729
            /rpt_family="Simple_repeat"
exon         18413..18530
            /gene="rac1"
            /number=3
            18531..22803
Intron       /gene="rac1"
            /number=3
            19287..19436
repeat_region /rpt_family="MIR"
            21161..21184
repeat_region /rpt_family="Simple_repeat"
            21185..21351
repeat_region /rpt_family="Alusq"
            21365..21403
repeat_region /rpt_family="Alu"
            21364..21365
misc_difference 21364..21365

```

/note="2345 nt insertion sequence in RPCI 11-derived BAC"

Query Match 58.5%; Score 203.6; DB 9; Length 28567;

Best Local Similarity 78.8%; Pred. No. 1.9e-41;

Matches 290; Conservative 0; Mismatches 54; Indels 24; Gaps 3;

```

QY      4 AACATCAGTGTCTTGCCAGAGTACCAACACTGTGATGTGATGATGCCGCCCTCTAGA 63
         |||||||
Db      27782 AACATCAGTGTCTTGCCAGAGTACCGACACTGTGACTGACCAATACTGA-CCCTCTTGA 27840

QY      64 CCTCACCCAGCGGACACATGCTCCGCTA-----CCTTGGCT 102
         |||||
Db      27841 CCTGCCACGCGGACACACGCTCCTGTAGTGCCTTGGCTTATGATGTTCCCTTGGGT 27900

QY      103 CTGTGAGTTCGTGTCAA--GCGCTAGTGTAAACGCGTTCGTGACAACCTACTACTGG 160
         |||||||
Db      27901 CTGTGAGTTCGTGTAAACTGTGTAGTGTGAGAGATGTTGTGACAACCTACTACTGG 27960

QY      161 CAAGAACACAGTGTGGCCCTTTCGACCACTAGAACAACTTTTCAATGACAGTTGC 220
         |||||
Db      27961 CGAAGATACAGCGCTGGACCCCTTCAGCACTAGAACAGAAATTTTAAATGACAGTTGC 28020

QY      221 AGAATGTGGAGTGTTTTACATTGATCTTGTGCTAATGACAGTATGAGATGTTTGGCA 280
         |||||||
Db      28021 AGAATGTGGAGTGTTTTACATTGATCTTGTGCTAATGACAGTATGAGATGTTTGGCA 28080

QY      281 TGTATGACTTAATTAATCTTGAATCATATAAAAAAAAAAAAAAAAAATGCTTTGGAACCTG 340
         |||||||
Db      28081 TGTATGACTTAATTAATCTTGAATCATATGACTGTGAATATGCGTGTTTTGAAGACTTG 28140

QY      341 AAAAAAAAA 348
         |||||
Db      28141 ATGACAA 28148

```

```

RESULT 11
AC009412      212827 bp   DNA       linear   PRI 07-NOV-2001
LOCUS         Homo sapiens BAC clone RP11-425P5 from 7, complete sequence.
DEFINITION    AC009412
AC009412.6   GI:14190769
VERSION       HTG.
KEYWORDS      Homo sapiens (human)
SOURCE        Homo sapiens
ORGANISM      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE     1 (bases 1 to 212827)
AUTHORS       Sulston, J.E. and Waterston, R.
TITLE         Toward a complete human genome sequence
JOURNAL       Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE       99063792
PUBMED        9847074
REFERENCE     2 (bases 1 to 212827)
AUTHORS       Hou, S., Maupin, R., Haakenson, W., Gregory, S. and Belter, E.
TITLE         The sequence of Homo sapiens BAC clone RP11-425P5
JOURNAL       Unpublished
REFERENCE     3 (bases 1 to 212827)
AUTHORS       Waterston, R.H.
TITLE         Direct Submission
JOURNAL       Submitted (21-AUG-1999) Genome Sequencing Center, Washington
              University School of Medicine, 4444 Forest Park Parkway, St. Louis,
              MO 63108, USA
REFERENCE     4 (bases 1 to 212827)
AUTHORS       Waterston, R.H.
TITLE         Direct Submission
JOURNAL       Submitted (23-MAY-2001) Genome Sequencing Center, Washington
              University School of Medicine, 4444 Forest Park Parkway, St. Louis,
              MO 63108, USA
REFERENCE     5 (bases 1 to 212827)
AUTHORS       Waterston, R.
TITLE         Direct Submission
JOURNAL       Submitted (07-NOV-2001) Department of Genetics, Washington
              University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
              On May 23, 2001 this sequence version replaced g1.13431187.
COMMENT       ----- Genome Center
              Center: Washington University Genome Sequencing Center
              Center code: WUGSC
              Web site: http://genome.wustl.edu/gsc
              Contact: sapiens@wustl.wustl.edu
              ----- Summary Statistics
              -----
              Center project name: H_NH0425P05
  
```

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgri.nih.gov/DIR/CTB/CHR7>, send <mailto:egreen@nhgri.nih.gov>, or see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

The RPc1-11 human BAC library was made from the blood of one male

FEATURES

donor, as described by Osogawa, K., Moon, P. Y., Zhao, B., Frengen, E., Tateo, M., Catanesse, J. J., and de Jong, P. J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. *Genomics* 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>) VECTOR: pBAC3.6

NEIGHBORING SOURCE INFORMATION:
The clone sequenced to the right is CTD-2195F2, 200 bp overlap. Actual start of this clone is at base position 1 of RP11-425P5; actual end is at base position 212633 of RP11-425P5.

Location/Qualifiers

```

1. 212827
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="7"
/map="7"
/clone="RP11-425P5"
/clone_lib="RPc1-11"
repeat_region 1623..1696
/rpt_family="MIR"
repeat_region 1698..2006
/rpt_family="Alu"
repeat_region 2024..2337
/rpt_family="Alu"
repeat_region 3006..3093
/rpt_family="L1"
repeat_region 3453..3552
/rpt_family="L2"
repeat_region 4924..5046
/rpt_family="ERV1"
misc_feature 5318..5693
/note="similar to EST BF96284 (NTD:912402607)"
repeat_region 5580..5845
/rpt_family="Alu"
repeat_region 5925..6277
/rpt_family="MALR"
misc_feature 6842..7041
/note="similar to EST BF369507 (NTD:911313532)"
repeat_region 7364..7661
/rpt_family="Alu"
repeat_region 8029..8048
/rpt_family="(CAAAA)n"
repeat_region 8379..8678
/rpt_family="Alu"
misc_feature 8560..8762
/note="similar to EST AA699398 (NTD:92702592) z140a03.s1"
repeat_region 8720..9031
/rpt_family="Alu"
misc_feature 9083..9560
/note="similar to EST AA633603 (NTD:92556817) ae66b02.s1"
repeat_region 9598..9899
/rpt_family="Alu"
misc_feature 9871..10098
/note="similar to EST AA699398 (NTD:92702592) z140a03.s1"
repeat_region 10746..11044
/rpt_family="L1"
repeat_region 11075..11101
/rpt_family="AT-rich"
repeat_region 11102..11372
/rpt_family="Alu"
repeat_region 11411..11465
/rpt_family="GA-rich"
repeat_region 11552..11644
/rpt_family="L2"
repeat_region 11685..12193
/rpt_family="L1"
repeat_region 12194..12486
/rpt_family="Alu"
repeat_region 12487..13003
/rpt_family="L1"
  
```

repeat_region	13004. .13305	/rpt_family="Alu"	
repeat_region	13306. .13501	/rpt_family="L1"	
repeat_region	13503. .13892	/rpt_family="L1"	
misc_feature	13893. .14224	/note="similar to	EST BE061102 (NID:g8405752)"
repeat_region	14244. .14544	/rpt_family="Alu"	
repeat_region	14546. .14828	/rpt_family="Alu"	
repeat_region	15091. .15189	/rpt_family="MIR"	
repeat_region	15672. .16449	/rpt_family="ACHOB0"	
repeat_region	16640. .16844	/rpt_family="Alu"	
repeat_region	16917. .16959	/rpt_family="ACHOB0"	
repeat_region	16960. .17287	/rpt_family="Alu"	
repeat_region	17288. .17572	/rpt_family="ACHOB0"	
misc_feature	17610. .17938	/note="similar to	EST BF996427 (NID:g12402750)"
repeat_region	17939. .18068	/rpt_family="Alu"	
repeat_region	18251. .18537	/rpt_family="Alu"	
misc_feature	18485. .18884	/note="similar to	EST BE062347 (NID:g8406997)"
misc_feature	18661. .18809	/note="similar to	EST BG186544 (NID:g13708231)"
repeat_region	18957. .19160	/rpt_family="L1"	
repeat_region	19314. .19528	/rpt_family="L1"	
repeat_region	19819. .20135	/rpt_family="Alu"	
repeat_region	21101. .21269	/rpt_family="L1"	
repeat_region	21882. .22203	/rpt_family="L1"	
misc_feature	22429. .22797	/note="similar to	EST BF993478 (NID:g12399801)"
repeat_region	22717. .23007	/rpt_family="L1"	
misc_feature	22993. .23191	/note="similar to	EST BF767044 (NID:g12115035)"
repeat_region	23167. .23471	/rpt_family="Alu"	
misc_feature	23511. .23930	/note="similar to	EST AN105611 (NID:g6076346) xda9g02.x1"
repeat_region	25387. .25677	/rpt_family="Alu"	

Query Match	58.5%	Score 203.6	DB 9	Length 212827
Best Local Similarity	78.8%	Pred. No. 2e-41		
Matches	290	Conservative	0	Mismatches 54; Indels 24; Gaps 3;
OY	4	AACATCACGCTTCCAGAGTACCAACATCTCATGTGATGATCCGCCCTCTAATA	63	
Db	175810	AACATCACGCTTCTCCAGATTACGCACACTGTCACTTGAACCACTGA-CGCTCTTAA	175868	
OY	64	CCTCACCACGGGAGACATGCTTCCGGTA-----CCTTTGGGT	102	
Db	175869	CCTCCCCACGGGAGACAGCGCTCTCTGTAGTCCTTGGCTTATGATGATCTCTTTGGGT	175928	
OY	103	CTGTAGAGTCTGTCA--GCGCTAGTGGTAAAGCGCGTCTGTACAACTTAATCACTG	160	
Db	175929	CTGTAGAGTCTGTAAACCTGTCTATGTGCTGACGATGTTCTGTACAACTTAATCACTG	175988	

QY	161	AAAGACACAGCTGTTGGGCGCTTTGCGACACATGACAAACCTTTTTCATTTACATGACAGTTCG	220
Db	175989	CGAGATACACACGCGTGGACCCCTTCGCCACATCAACAGATATTTTAAATTACAGTTGC	176048
QY	221	AGAAATTGTGGAGGTGTTTTTACATTCATGCTTTTGTCTAATGACAGTTGACAGTATGTTTGGCA	280
Db	176049	AGAAATTGTGGAGGTGTTTTTACATTCATGCTTTTGTCTAATGACAGTTGACAGTATGTTTGGCA	176108
QY	281	TGTTATGACTTAAATTAATCTCTGAAATCATATTAATAAAAAAAAAAATGCTTTGGAACTTG	340
Db	176109	TGTTATGACTTAAATTAATTAATCTCTGAAATCATATGAGCTGTAATGCTGTTGTTTGGAGACTTG	176168
QY	341	AAAAAAAA 348	
Db	176169	ATGAACAAC 176176	
RESULT 12			
AK054993		2315 bp mRNA linear PRI 01-AUG-2002	
LOCUS			
DEFINITION		Homo sapiens CDNA FLJ30431 fls, clone BRACE2008968, highly similar	
ACCESSION		AK054993.1 GI:16549633	
VERSION		AK054993.1	
KEYWORDS		oligo capping; fls (full insert sequence).	
SOURCE		Homo sapiens (human)	
ORGANISM		Homo sapiens	
REFERENCE		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	
AUTHORS		1 Tashiro,H., Yamazaki,M., Matanabe,K., Kumagai,A., Itakura,S., Furuzumi,T., Fujimori,Y., Komiyama,M., Sugiyama,T., Ito,R., Otsuki,T., Sato,H., Wakamatsu,A., Ishii,S., Yamamoto,J., Isono,Y., Kawai-Hio,Y., Saito,K., Nishikawa,T., Kimura,K., Yamashita,H., Matsuo,K., Nakamura,Y., Sekine,M., Kikuchi,H., Kanda,K., Matsushima,M., Murekawa,K., Kanehori,K., Takahashi-Fujii,A., Oshima,A., Sugiyama,A., Kawakami,B., Suzuki,Y., Sugano,S., Nagahatai,K., Masuno,Y., Nagai,T. and Sugiyama,T.	
TITLE		NEDO human cDNA sequencing project	
JOURNAL		Unpublished	
REFERENCE		2 (bases 1 to 2315)	
AUTHORS		Isogai,T., Otsuki,T. and Sugiyama,T.	
TITLE		Direct Submission	
JOURNAL		Submitted (24-OCT-2001) Takao Isogai, Helix Research Institute, Genomics Laboratory; 1352-3 Yana, Kiseazuru, Chiba 292-0812, Japan (E-mail:genomics@helix.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)	
COMMENT		NEDO human cDNA sequencing project supported by Ministry Of Economy, Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology (RAB); cDNA library construction: Helix Research Institute (HRI) (supported by Japan Key Technology Center etc.); 5'- and 3'-end one pass sequencing: RAB, HRI, and Biotechnology Center, National Institute of Technology and Evaluation; clone selection for full insert sequencing: RAB and HRI.	
FEATURES			
source		Location/Qualifiers	
		1..2315	
		/organism="Homo sapiens"	
		/mol_type="mRNA"	
		/db_xref="taxon:9606"	
		/clone="BRACE2008968"	
		/tissue_type="cerebellum"	
		/clone_id="BRACE2"	
		/note="cloning vector: pME185FL3"	
BASE COUNT		603 a 531 c 501 g 680 t	
ORIGIN			
Query Match		55.3%; Score 192.4; DB 9; Length 2315;	
Best Local Similarity		81.7%; Pred. No. 1.2e-38;	
Matches		268; Conservative 0; Mismatches 36; Indels 24; Gaps 3;	
QY	4	AACATCACTGCTTGGCCAGAGTACCAACACTGTCATGTGATTGATGCCGCCCTCTYAGA	63
Db	1989	AACATCACTGCTTGGCCAGATTAACGCACTGCTGATACCAATATCTGA-CCTCTTTTA	2047

REFERENCE	AUTHORS	TITLE
1 (Pages 1 to 1232)	Ishida, Y., Hadano, S., Nagayama, T., Tomiyasu, H., Wakasa, K. and Ikeda, J.	Isolation and characterization of 21 novel expressed DNA sequences from the distal region of human chromosome 4p

JOURNAL Genomice 22 (2), 302-312 (1994)
MEDLINE 95104839
PubMed 7545969
REFERENCE 2 (bases 1 to 1232)
AUTHORS Ishida,Y.
TITLE Direct Submission
JOURNAL Submitted (11-NOV-1993) Yoshikazu Ishida, Ikeda Genosphere
Project/ERATO/JRDC, Tokai University School of Medicine, Bohseidai,
Isehara, Kanagawa 259-1193, Japan
(E-mail:shinjien@med.u-tokai.ac.jp, Tel:81-463-91-5095,
Fax:81-463-91-4993)
On Mar 25, 1994 this sequence version replaced g1:434743.
Submitted (11-Nov-1993) to DDBJ by:
Yoshikazu Ishida
Ikeda Genosphere Project/ERATO/JRDC
Tokai University School of Medicine
Bohseidai, Isehara
Kanagawa 259-11
Japan
Phone: 0463-91-4056
Fax: 0463-91-4110.
Location/Qualifiers
1. .1232
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/chromosome="4"
/clone="PO2ST9"
/issue_type="brain striatum"
POLYA_signal 1214. .1219
POLYA_site 1232
BASE COUNT 329 a 240 c 251 g 411 t 1 others
ORIGIN

Query Match 53.9%; Score 187.4; DB 9; Length 1232;
Best Local Similarity 80.9%; Pred. No. 2.2e-37;
Matches 263; Conservative 1; Mismatches 37; Indels 24; Gaps 3;

QY 4 AACATCACTGCTTCCAGAGTACCAACACTGTGATGATGCGCCCTCTAGA 63
|||||
Db 905 AACATCACTGCTTCCAGAGTACCAACACTGTGATGATGCGCCCTCTAGA 963
64 CCTCACCACGCGGACACATGCTTCGGTA-----CCTTGGGT 102
|||||
Db 964 CCTCACCACGCGGACACACGCTCCTGTAGTGCCTTGCCTATTGATGTTCTTGGGT 1023
103 CTGTAGAGTCTGTCAA--GCGCTAGTCTTAACGCGTCTGTACAACTTAAGTCACTGG 160
|||||
Db 1024 CTGTAGAGTCTGTAAACTGTCTAGTGTGACGATGTTCTGTACAACTTAAGTCACTGG 1083
QY 161 CAAGAACACAGTGTGGGCTTGCACACTAGAACAACTTTTCAATTGACAGTTGC 220
|||||
Db 1084 CGAATAATACAGCGTGGACCTTCAGCACTACACGAATTTTAAATTGACAGTTGC 1143
QY 221 AGAATTGTGAGTGTCTTATCATTTGATCTTGTCTAATGACAGTATGATGTTTGA 280
|||||
Db 1144 AGAATTGTGAGTGTCTTATCATTTGATCTTGTCTAATGACAGTATGATGTTTGA 1203
QY 281 TGTATGACTTAATAATCCTGAAT 305
|||||
Db 1204 TGTATGACTTAATAATCCTGGAM 1228

Search completed: August 26, 2003, 02:30:08
Job time : 1288.94 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 25, 2003, 19:52:32 ; Search time 96.0274 Seconds
(Without alignments)
9782.674 Million cell updates/sec

Title: US-09-717-321A-17

Perfect score: 348
Sequence: 1 tgaacatcactgtcttgc.....cttggaactgaaaaaaa 348

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

N_Geneseq_19Jun03:*

1: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*
2: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*
3: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*
4: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*
5: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*
6: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*
7: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*
8: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*
9: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*
10: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*
11: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*
12: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*
13: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*
14: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*
15: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*
16: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*
17: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*
18: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*
19: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*
20: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*
21: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*
22: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*
23: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*
24: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*
25: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	348	100.0	348	22	AAH22399
2	348	100.0	1266	22	AAH22396
3	209.4	60.2	2051	21	AAF18236
4	203.6	58.5	28567	25	ABT17030
5	187.8	54.0	352	22	AAH22400
6	187.4	53.9	1232	24	ABK83623
7	187.4	53.9	1232	24	ABN95134
8	187.4	53.9	1318	23	AAS65569

C	9	163.6	47.0	353	25	ABX46267
C	10	160.6	46.1	447	21	AAA89693
C	11	145.2	41.7	422	21	AAA89694
C	12	123	35.3	123	22	AAH22395
C	13	116.4	32.4	3740	22	AAH07296
C	14	113	32.5	1017	22	AAH22397
C	15	95.4	27.4	201	25	ABX39192
C	16	63.8	18.3	68	16	AAT22394
C	17	60.8	17.5	64	25	ABZ78455
C	18	60.8	17.5	64	25	ABZ09002
C	19	52.4	15.1	162	24	ABK45998
C	20	52.4	15.1	605	21	AAH16478
C	21	52.4	15.1	1022	22	AAH22398
C	22	50.8	14.6	605	24	ABL38065
C	23	50	14.4	50	24	ABZ01605
C	24	43.2	12.4	424	25	ABX46053
C	25	43	12.4	18660	22	AAK79108
C	26	43	12.4	38844	24	AAH40255
C	27	42.2	12.1	640681	24	ABA92787
C	28	42	12.1	15261	22	AAH07498
C	29	41.8	12.0	9817	24	ABL33368
C	30	41.4	11.9	6464	24	ABL32514
C	31	40.8	11.7	5852	12	AAQ11710
C	32	40.4	11.6	501	20	AAH97621
C	33	40.2	11.6	185371	24	ABT10718
C	34	40	11.5	534	23	ABV54457
C	35	39.6	11.4	1494	22	AAK86444
C	36	38.6	11.4	2101	22	AAK86445
C	37	38.6	11.4	17918	24	AAH61419
C	38	38.4	11.3	6059	22	ABL32813
C	39	39.4	11.3	6409	22	AAH46495
C	40	39.2	11.3	277	25	ABX47508
C	41	39.2	11.3	375	25	ABX49849
C	42	39.2	11.3	859	17	AAT32572
C	43	38.2	11.3	2039	17	AAT32569
C	44	39.2	11.3	6621	22	ABL32913
C	45	39.2	11.3	18671	22	AAK90763

ALIGNMENTS

RESULT 1	AAH22399	standard; DNA; 348 BP.
ID	AAH22399	
AC	AAH22399;	
XX		
DT	22-AUG-2001	(first entry)
XX		
DE	Human rac1 gene related nucleotide sequence #3.	
XX		
KW	Identification; toxic; hepatotoxic; differential gene expression;	
KW	NSAID; non-steroidal antiinflammatory drug; ds.	
XX		
OS	Homo sapiens.	
XX		
PN	WO200138579-A2.	
XX		
PD	31-MAY-2001.	
XX		
PF	21-NOV-2000; 2000MO-US32049.	
XX		
PR	22-NOV-1999; 99US-0166923.	
PR	18-FEB-2000; 2000US-0185311.	
PR	20-NOV-2000; 2000US-0717321.	
PA	(CURA-) CURAGEN CORP.	
XX		
PI	Gould-Rothberg BE, Dipippo VA, Ramsehl TM, Gerwehn RW;	
XX		
DR	WPI, 2001-355948/37.	
XX		

Bovine EST associa
Mouse Exo103 nucle
Mouse Rab2 nucleot
Human rac1 genomic
Human reproductive
Human rac1 gene re
Bovine EST associa
Human gene signatu
Tumour suppression
Human oligonucleot
cDNA encoding colo
Human colon cancer
Human rac1 gene re
Human colon tumour
Human leukocyte ge
Bovine EST associa
Human immune/haema
Genomic DNA encodl
Buchera sp. genom
Human reproductive
Human immune syste
Dictyostelium plas
Extended human sec
Human breast cance
Human prostate exp
Human immune/haema
Human gene regulat
Human immune syste
Tumour suppressor
Bovine EST associa
Bovine EST associa
Microspore-specifi
Rapeseed microspor
Human immune syste
Human digestive sy

PT	Screening hepatotoxic agent comprises contacting test cell population
PT	expressing RISKMARKER or INJURYMARKER with agent, comparing expression
PT	with reference population and identifying difference in expression
PT	levels
XX	-
PS	Disclosure; page 8-9; 76pp; English.
XX	
CC	The present invention describes a method of screening a test agent for
CC	hepatotoxicity. The method comprises: (a) providing a test cell
CC	population comprising a cell capable of expressing one or more nucleic
CC	acid sequences selected from the group consisting of RISKMARKER 1-8
CC	and INJURYMARKER 1-10; (b) contacting the test cell population with a
CC	test agent; (c) measuring expression of one or more of the nucleic
CC	acid sequences in the test cell population; (d) comparing the
CC	expression of the nucleic acid sequence in the test cell population to
CC	the expression of the nucleic acid sequence in a reference cell
CC	population comprising at least one cell whose exposure status to a
CC	hepatotoxic agent is known; and (e) identifying a difference in
CC	expression levels of the RISKMARKER or INJURYMARKER sequences, if
CC	present, in the test cell population and reference cell population.
CC	The method is useful for identifying a hepatotoxic agent. The present
CC	sequence is given in the exemplification of the present invention.
XX	
SO	Sequence 348 BP; 103 A; 76 C; 67 G; 102 T; 0 other:
Query Match	100.0%; Score 348; DB 22; Length 348;
Best Local Similarity	100.0%; Pred. No. 2.5e-77;
Matches 348; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
OY	1 TGAACATCAGCTGCTTGGCCAGAGTACCAACAGCTGCATGTGATGATGCCGCCCTCT 60
DB	1 TGAACATCAGCTGCTTGGCCAGAGTACCAACAGCTGCATGTGATGATGCCGCCCTCT 60
OY	61 AGACCTCACCGACGGGACACATGCTTCGGTACTTTGGGCTGTGAGTTCTGTCAAG 120
DB	61 AGACCTCACCGACGGGACACATGCTTCGGTACTTTGGGCTGTGAGTTCTGTCAAG 120
OY	121 CGCTAGTCTTAAGCGCGCTTCTGTACAACCTAATCACTGACACGCAAGACAGTGGGCC 180
DB	121 CGCTAGTCTTAAGCGCGCTTCTGTACAACCTAATCACTGACACAGAGTGGGGCC 180
OY	181 TTTGCACCACTAGAACAACTTTTTCAAATTGACAGTTGCAGAAATTGGAGTGTTTTGA 240
DB	181 TTTGCACCACTAGAACAACTTTTTCAAATTGACAGTTGCAGAAATTGGAGTGTTTTGA 240
OY	241 CATTATCTTTTGGTAATGCACTTAGCAGTATGTTTGCATGTATGACTTAATAATCCT 300
DB	241 CATTATCTTTTGGTAATGCACTTAGCAGTATGTTTGCATGTATGACTTAATAATCCT 300
OY	301 TGAATCATTAATAAAAAAAAAAAAAAATGCTTTGGAACTTGAAGAAAAA 348
DB	301 TGAATCATTAATAAAAAAAAAAAAAAATGCTTTGGAACTTGAAGAAAAA 348
AAH22396/C	
AAH22396	standard; DNA; 1266 BP.
AC	
AC	AAH22396;
XX	
XX	22-AUG-2001 (first entry)
XX	
XX	Human rac1 contig SEQ ID NO:2.
XX	
XX	Identification; toxic; hepatotoxic; differential gene expression;
XX	NSAID; non-steroidal antiinflammatory drug; ds.
XX	
XX	Homo sapiens.
XX	
XX	WO200138579-A2.
XX	
XX	31-MAY-2001.
XX	

FE	21-NOV-2000;	2000OWO-US32049.
XX		
PR	22-NOV-1999;	99US-0166923.
PR	18-FEB-2000;	2000OUS-0183531.
PR	20-NOV-2000;	2000OUS-0717321.
XX		
PA	(CURAGEN CORP.	
XX		
PI	Gould-Rothberg BE, D'Alipho VA, Ramseh TM, Gerwein RW;	
XX		
DR	WPI; 2001-355948/37.	
XX		
PT	Screening hepatotoxic agent comprises contacting test cell population	
PT	expressing RISKMARKER or INURYMARKER with agent, comparing expression	
PT	with reference population and identifying difference in expression	
PT	levels	
XX		
PS	Disclosure; Page 7; 76pp; English.	
XX		
CC	The present invention describes a method of screening a test agent for	
CC	hepatotoxicity. The method comprises: (a) providing a test cell	
CC	population comprising a cell capable of expressing one or more nucleic	
CC	acid sequences selected from the group consisting of RISKMARKER 1-8	
CC	and INURYMARKER 1-10; (b) contacting the test cell population with a	
CC	test agent; (c) measuring expression of one or more of the nucleic	
CC	acid sequences in the test cell population; (d) comprising the	
CC	expression of the nucleic acid sequence in the test cell population to	
CC	the expression of the nucleic acid sequence in an reference cell	
CC	population comprising at least one cell whose exposure status to a	
CC	hepatotoxic agent is known; and (e) identifying a difference in	
CC	expression levels of the RISKMARKER or INURYMARKER sequences, if	
CC	present, in the test cell population and reference cell population.	
CC	The method is useful for identifying a hepatotoxic agent. The present	
CC	sequence is given in the exemplification of the present invention.	
XX		
SO	Sequence 1266 BP; 385 A; 258 C; 285 G; 338 T; 0 other;	
	Query Match	100.0%; Score 348; DB 22; Length 1266;
	Best Local Similarity	100.0%; Pred. No. 3.5e-77;
	Matches 348; Conservative	100.0%; Mismatches 0; Indels 0; Gaps 0;
OY	1	TGAACAACACTGCTGCTTCCAGAGACCAACACTGTCAATGATGATGCGCCCTCT 60
DB	357	TGAAACAATCACTGCTTCCAGAGACCAACACTGTCAATGATGATGCGCCCTCT 298
OY	61	AGACCTCACCCAGCGGACACATGCTCCGGTACCTTTGGGTCTGTGAAGTTCTGCAAG 120
DB	297	AGACCTCACCCAGCGGACACATGCTCCGGTACCTTTGGGTCTGTGAAGTTCTGCAAG 238
OY	121	CGGTAGTCTTAACGGCGTCTGTCAACCTTAACCTACACCTGCGCAAGAACACAGTTGGGCC 180
DB	237	CGGTAGTCTTAACGGCGTCTGTCAACCTTAACCTACACCTGCGCAAGAACACAGTTGGGCC 178
OY	181	TTTTCGACCACTAGAACAACTTTTTTCAATTGACAGTGTGCAAGAAATTGGAGTGTTTTAA 240
DB	177	TTTTCGACCACTAGAACAACTTTTTTCAATTGACAGTGTGCAAGAAATTGGAGTGTTTTAA 118
OY	241	CATTGATTTTTTGTCTAATGCAATGCAATGTTTTCATGATGACTTAATAATCTT 300
DB	117	CATTGATTTTTTGTCTAATGCAATGCAATGTTTTCATGATGACTTAATAATCTT 58
OY	301	TGAATCAATAAAAAAAAAAAAAATGCTTTGGAACTTGAAGAAAAA 348
DB	57	TGAATCAATAAAAAAAAAAAAAATGCTTTGGAACTTGAAGAAAAA 10
	RESULT 3	
	AAFI8236	
ID	AAFI8236 standard; DNA; 2051 BP.	
XX		
AC	AAFI8236;	
XX		
DT	14-MAR-2001 (first entry)	

XX Lung cancer associated polynucleotide sequence SEQ ID 255.
 DE Human; lung cancer associated protein; neuroprotective; cytostatic;
 KW cardiocactive; immunomodulatory; muscular active; vulnerary;
 KW gastrointestinal; nephrotoxic; antiinfective; gynecological;
 KW antibacterial; diagnosis; neural disorder; immune disorder; reproductive;
 KW proliferative disorder; wound healing; infectious disease; ds.
 XX
 OS Homo sapiens.
 XX
 PN WO20005180-A2.
 PD 21-SEP-2000.
 XX
 PF 08-MAR-2000; 2000WO-US05918.
 XX
 PR 12-MAR-1999; 99US-0124270.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 PA (ROSE/) ROSEN C A.
 XX
 PI Ruben SM;
 XX
 DR WPI: 2000-587514/55.
 DR P-PSDB: AAB58360.
 XX
 PT Lung cancer associated gene sequences, referred to as lung cancer
 PT antigens, useful for treatment, prevention, and diagnosis of disorders
 PT such as lung cancer -
 XX
 XX Claim 1; Page 716-717; 1425pp; English.
 PS
 XX Polynucleotide sequences AAF17982 - AAF18424 encode human lung cancer
 CC associated proteins represented in AAB58106 - AAB58548. Lung cancer
 CC associated proteins and polynucleotide sequences, their agonists, and
 CC antagonists may have neuroprotective; cytostatic; cardioactive;
 CC immunomodulatory; muscular active general; vulnerary; gastrointestinal
 CC general; nephrotoxic; antiinfective; gynecological; or antibacterial
 CC activity. The invention also includes antibodies specific for the
 CC protein or polynucleotide sequences. The lung cancer associated
 CC polynucleotide sequences may be used for detection of lung cancer,
 CC chromosome identification, as chromosome markers, and for numerous other
 CC diagnostic or research purposes. The proteins may be used to treat
 CC disorders such as neural, immune, muscular, reproductive,
 CC gastrointestinal, pulmonary, cardiovascular, renal, and proliferative
 CC disorders. The proteins may also be used in the treatment of wounds and
 CC infectious diseases. Polynucleotide sequences AAF18425 - AAF18433 and
 CC peptide AAB58549 are used in the course of the invention for the
 CC identification and characterization of the polynucleotide and protein
 CC sequences.
 CC
 XX Sequence 2051 BP; 570 A; 430 C; 433 G; 612 T; 6 other:
 SQ
 Query Match 60.2%; Score 209.4; DB 21; Length 2051;
 Best Local Similarity 82.6%; Pred. No. 1.4e-42;
 Matches 285; Conservative 0; Mismatches 36; Indels 24; Gaps 3;
 OY 4 AACATCACTGTCTTCCAGAGTACCAACATGTCTATGATGATGCGCCCTCTAGA 63
 DB 1666 AACATCACTGTCTTCCAGATTTACCGACAGCTGTCTAGTACCAATATGCA -CCCTCTTGA 1724
 OY 64 CCTGACCCAGCGGACACATCTTCCGGTA-----CCTTTGGGT 102
 DB 1725 CCTGCGCCAGCGGACACACGCCCTCTAGTCCGTTGCCATTTGATGTTCTTTGGGT 1784
 OY 103 CTGTGAGTTCTGTCA--GGCTAGTGTCTAAAGCCGTTCTGTACACCTTACTGCTG 160
 DB 1785 CTGTGAGTTCTGTAACTGTGTCTAGTGTCTGATGATGTTCTGTACAACTTAACTGCTG 1844
 OY 161 CAGAGACAGAGTGTGGGCGTTTGACACAGTAAAGAACTTTTTCATTTGACAGTTGC 220
 DB 1845 CGAGATACAGCGTGGGACCCCTTACGCCATACACAGAAATTTTAAATGACAGTTGC 1904

OY 221 AGAATTTGAGAGTGTCTTTTACATGATCTTTTGTCTAATGACATGACATGATTTGCA 280
 DB 1905 AGAATTTGAGAGTGTCTTTTACATGATCTTTTGTCTAATGACATGACATGATTTGCA 1964
 OY 281 TGTATGACTTAAATTAATCCTTGATCATATATAAAAAAAAAAAAAA 325
 DB 1965 TGTATGACTTAAATTAATCCTTGATCATATATAAAAAAAAAAAAAA 2009
 RESULT 4
 ABE17030
 ID ABE17030 standard; DNA; 28567 BP.
 XX
 AC ABE17030;
 XX
 DT 03-APR-2003 (first entry)
 XX
 DE Human MP21 gene Rac1 SEQ ID No 4.
 XX
 KW Cytostatic; p21 pathway modulating agent; cancer; angiogenic; apoptotic;
 KW cell proliferation disorder; MP21; gene; ds.
 XX
 OS Homo sapiens.
 XX
 PN WO2003006990-A1.
 XX
 PD 23-JAN-2003.
 XX
 PF 10-JUL-2002; 2002WO-US21549.
 XX
 PR 12-JUL-2001; 2001US-305017P.
 PR 10-OCT-2001; 2001US-328491P.
 PR 15-FEB-2002; 2002US-357452P.
 XX
 PA (EXEL-) EXELIXIS INC.
 PI
 DR Friedman L, Plowman GD, Belvin M, Li D, Funke RP;
 DR WPI: 2003-221779/21.
 DR P-PSDB: ABE19756.
 XX
 XX Identifying candidate p21 pathway modulator, by contacting an assay
 PT system having modifiers of p21 polypeptide or gene with a test agent to
 PT provide a reference activity in system and detecting test agent-biased
 PT activity -
 XX
 PS Examples; Page 56-72; 199pp; English.
 XX
 CC The invention relates to a novel method for identifying a candidate p21
 CC pathway modulating agent. The novel method comprises contacting an assay
 CC system, comprising a purified MP21 polypeptide (modifier of p21) or
 CC nucleic acid, with a test agent under conditions, so that but for the
 CC presence of a test agent, the assay system provides a reference activity
 CC and detection of test agent-biased activity of the assay system. The
 CC novel method of the invention is useful for identifying a candidate p21
 CC pathway modulating agent. The invention also includes a method for
 CC modulating the p21 pathway of a cell, and a method for diagnosing a
 CC disease e.g. cancer in a patient. The identified modulators are useful in
 CC diagnosis, therapy and pharmaceutical development. The modulators are
 CC useful in a variety of diagnostic and therapeutic applications including
 CC angiogenic, apoptotic and cell proliferation disorders. This
 CC polynucleotide sequence represents a gene encoding an MP21 protein of the
 CC invention.
 CC
 XX Sequence 28567 BP; 6762 A; 6358 C; 6796 G; 8651 T; 0 other:
 SQ
 Query Match 58.5%; Score 203.6; DB 25; Length 28567;
 Best Local Similarity 78.8%; Pred. No. 7.5e-41;
 Matches 290; Conservative 0; Mismatches 54; Indels 24; Gaps 3;
 OY 4 AACATCACTGTCTTCCAGAGTACCAACATGTCTATGATGATGCGCCCTCTAGA 63
 DB 1666 AACATCACTGTCTTCCAGATTTACCGACAGCTGTCTAGTACCAATATGCA -CCCTCTTGA 1724

DB 27782 AACATCAGTCTCTGGCAGATTACGACACTGTCACTTGACCAATACTGA-CCCTCTTTA 27840
QY 64 CCTCACCACGGCGGACACATGCTTCGGTA-----CCTTGGGT 102
DB 27841 CTTCCGCCACGGGACACAGCCTCTGTAGCGTTTGCCATGTGATGTTCTTTGGGT 27900
QY 103 CTTGAGGTTCTCTCA--GCCCTAGTGTCAAGCCGTTCTGTACAACCTAATCTCATG 160
DB 27901 CTTGAGGTTCTCTCAAGTGTGCTGAGAGTGTCTGTACAACCTAATCTCATG 27960
QY 161 CAGAACACAGTGTGGGCTTTCGACCACTAGAACAACTTTTCAATGTGACAGTTC 220
DB 27961 CGAGATACAGCGTGGGACCTTCACCCACTACACAGATTTTAAATGTGACAGTTC 28020
QY 221 AGAATGTGAGTGTTTTACATGATCTTTTGTAAATGACAGTACAGTATGTTTGA 280
DB 28021 AGAATGTGAGTGTTTTACATGATCTTTTGTAAATGACATATATGTTTGA 28080
QY 281 TGTATGACTTAATAATCCCTGTAATGATATAAAAAAATGCTTTTGAGACTTG 340
DB 28081 TGTATGACTTAATAATCCCTGTAATGATGAGTGTATATGTTTGAAGACTTG 28140
QY 341 AAAAAAA 348
DB 28141 ATGAAACA 28148
RESULT 5
AAH22400
ID AAH22400 standard; DNA; 352 BP.
XX
AC AAH22400;
XX
DT 22-AUG-2001 (first entry)
XX
DE Human rac1 gene related nucleotide sequence #4.
XX
KW Identification: toxic; hepatotoxic; differential gene expression;
XX
KW NSAIID; non-steroidal antiinflammatory drug; ds.
XX
OS Homo sapiens.
XX
PN WO200138579-A2.
XX
PD 31-MAY-2001.
XX
PE 21-NOV-2000; 2000MO-US32049.
XX
PR 22-NOV-1999; 99US-0166923.
PR 18-FEB-2000; 2000US-0183531.
PR 20-NOV-2000; 2000US-0717321.
XX
PA (CURA-) CURAGEN CORP.
XX
PI Gould-Rothberg BE, DiPippo VA, Ramseh TM, Gerweln RW;
XX
DR WPI; 2001-355948/37.
XX
PT Screening hepatotoxic agent comprises contacting test cell population
PT with reference RISKMARKER or INJURYMARKER with agent, comparing expression
PT levels -
XX
PS Disclosure; Page 8-9; 76pp; English.
XX
CC The present invention describes a method of screening a test agent for
CC hepatotoxicity. The method comprises: (a) providing a test cell
CC population comprising a cell capable of expressing one or more nucleic
CC acid sequences selected from the group consisting of RISKMARKER 1-8
CC and INJURYMARKER 1-10; (b) contacting the test cell population with a
CC test agent; (c) measuring expression of one or more of the nucleic
CC acid sequences in the test cell population; (d) comparing the
CC expression of the nucleic acid sequence in the test cell population to

CC the expression of the nucleic acid sequence in a reference cell
CC population comprising at least one cell whose exposure status to a
CC hepatotoxic agent is known; and (e) identifying a difference in
CC expression levels of the RISKMARKER or INJURYMARKER sequences. If
CC present, in the test cell population and reference cell population.
CC The method is useful for identifying a hepatotoxic agent. The present
CC sequence is given in the exemplification of the present invention.
SQ Sequence 352 BP; 86 A; 77 C; 72 G; 117 T; 0 other;
Query Match 54.0%; Score 187.8; DB 22; Length 352;
Best Local Similarity 77.6%; Pred. No. 2,1e-37;
Matches 274; Conservative 0; Mismatches 57; Indels 22; Gaps 3;
QY 17 TGGCAGAGTACCAACACTGTATGATGATGCGCCGCCCTTAGACCTACACCGG 76
DB 1 TGGCAGATTTACGCAAGCTGTCACTTGACCAATAGCTG-ACCCCTTTACCTGCCGCGG 59
QY 77 GACACATGCTCCG-----GTACCTTTGGGCTCTGAGGTTCTGTC 117
DB 60 GACACCGCCCTCTGTGCTGCTTGCCTAATGATGTTCTTTGGGCTGTGAGGTTCTGTA 119
QY 118 AA--GGCGTATGCTGTAACGCCGTTCTGTACAACTTAACCTGCGAAGACAGTGT 175
DB 120 AACTGTGCTATGCTGCTGACGATGTTCTGTACAACTTAACCTGCGAAGATACAGCTG 179
QY 176 GGGCCTTTGACGACCTAGAACAACTTTTCAATGACAGTGTGAGAAATGTGAGTGT 235
DB 180 GGACCCCTTCAGCCACTACCAAGAAATTTTAAATGACAGTGTGAGAAATGTGAGTGT 239
QY 236 TTTTACATGATGCTTTTGTCAATGACAGTGTGATGTTTTCATGATGATGATTAATA 295
DB 240 TTTTACATGATGCTTTTGTCAATGACAGTGTGATGTTTTCATGATGATGATTAATA 299
QY 296 ATCCTTGAATCATTAATAAAAAAATGCTTTGGAACCTTGAATAAAAA 348
DB 300 ATCCTTGAATCATACGACCTGTATATCTGCTGTTTGAAGACTTGAATAA 352
RESULT 6
ABK83623
ID ABK83623 standard; cDNA; 1232 BP.
XX
AC ABK83623;
XX
DT 14-AUG-2002 (first entry)
XX
DE Human cDNA differentially expressed in granulocytic cells #194.
XX
KW Human; ss; granulocytic cell; DNA chip; bacterial infection;
KW viral infection; parasitic infection; protozoal infection;
KW fungal infection; sterile inflammatory disease; psoriasis;
KW rheumatoid arthritis; glomerulonephritis; asthma; thrombosis;
KW cardiac reperfusion injury; renal reperfusion injury; ARDS;
KW adult respiratory distress syndrome; inflammatory bowel disease;
KW Crohn's disease; ulcerative colitis; periodontal disease;
KW granulocyte activation; chronic inflammation; allergy.
XX
OS Homo sapiens.
XX
PN WO200228899-A2.
XX
PD 11-APR-2002.
XX
PE 03-OCT-2001; 2001MO-US30821.
XX
PR 03-OCT-2000; 2000US-237189P.
XX
PA (GENE-) GENE LOGIC INC.
XX
PI Beazer-Barclay Y, Weissman SM, Yamaga S, Vockley J;
XX
DR WPI; 2002-435328/46.

XX Detecting granulocyte activation by detecting differential expression
PT of genes associated with granulocyte activation, which serves as
PT diagnostic markers that is useful for monitoring disease states and
PT drug toxicity -
XX
PS Claim 1; SEQ ID NO 194; 114pp; English.
XX
CC The invention relates to detecting (M1) granulocyte (GC) activation
CC (GCA), by detecting the level of expression of gene(s) (Gs) identified by
CC DNA chip analysis as given in the specification, and comparing
CC the expression level to an expression level in an unactivated
CC GC, where differential expression of Gs is indicative of GCA.
CC Also included are modulating (M2) GA by contacting GC with an agent
CC that alters the expression of at least one gene in Gs; (2) screening (M3)
CC for an agent capable of modulating GCA or an inflammation (especially
CC chronic) in a tissue, an allergic response in a subject, exposure of a
CC subject to a pathogen or sterile inflammatory disease using the
CC gene expression profile; (3) detecting (M4) an inflammation (especially
CC chronic) in a tissue, an allergic response in a subject, exposure of a
CC subject to a pathogen or sterile inflammatory disease, by detecting the
CC level of expression in a sample of the tissue of gene(s) from Gs, where
CC the level of expression of the gene is indicative of inflammation;
CC (4) treating (M5) an inflammation (especially chronic) or in a tissue,
CC an allergic response in a subject, exposure of a subject to a pathogen
CC or sterile inflammatory disease, by contacting a tissue having
CC inflammation with an agent that modulates the expression of gene(s)
CC from Gs in the tissue. M1 is useful for detecting GCA; M2 is useful for
CC modulating GA; M3 is useful for screening an agent capable of modulating
CC GCA preferably in an inflammation in a tissue; M4 is useful for
CC detecting an inflammation (especially chronic) in a tissue, an allergic
CC response in a subject, exposure of a subject to a pathogen or sterile
CC inflammatory disease (e.g. psoriasis, rheumatoid arthritis,
CC glomerulonephritis, asthma, thrombosis, cardiac reperfusion injury, renal
CC reperfusion injury, AIDS, adult respiratory distress syndrome,
CC inflammatory bowel disease, Crohn's disease, ulcerative colitis,
CC periodontal disease; also bacterial infection, viral infection,
CC parasitic infection, protozoal infection, fungal infection and M5 is
CC useful for treating one of the above conditions. The present
CC sequence represents a gene differentially expressed in granulocytes.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1232 BP; 329 A; 240 C; 251 G; 411 T; 1 other:
Query Match 53.9%; Score 187.4; DB 24; Length 1232;
Best Local Similarity 80.9%; Pred. No. 3.6e-37;
Matches 263; Conservative 1; Mismatches 37; Indels 24; Gaps 3;
OY 4 AACATCACTGCTCTCCAGAGTACCAACATGTCATGTGATGAGCCGCCCTCTGA 63
DB 905 AACATCACTGCTCTCCAGAGTACCAACATGTCATGTGATGAGCCGCCCTCTGA 963
OY 64 CCTCACCCAGCGGACACATGCTTCGGTA-----CCTTTGGGT 102
DB 964 CCTGCGCCAGCGGACACACGCTCTGATGCTGCTTGCCTATGTAGTGTCTTTGGGT 1023
OY 103 CTGTGAGGTCTGTGCA--GGGCTAGTGTCTAACGCCCTTCTGTACACCTTACACG 160
DB 1024 CTGTGAGGTCTGTGCA--GGGCTAGTGTCTAACGCCCTTCTGTACACCTTACACG 1083
OY 161 CAGAACACAGTGTGGGCTTTCGACACATAGAACATCTTTTCAATGTGACAGTTGC 220
DB 1084 CGAGAAATACAGCGTGGGACCTTCAGCCACATCAACAATTTTAAATGTGACAGTTGC 1143
OY 221 AGAATTTGAGAGTGTTTTACATGTGATCTTTTCTAATGACAGTATGTTTGA 280
DB 1144 AGAATTTGAGAGTGTTTTACATGTGATCTTTTCTAATGACATATGCAATATGTTTGA 1203
OY 281 TGTATGACTTAAATCTTGAAT 305
|||||

DB 1204 TGTATGACTTAAATCTTGAAT 1228
RESULT 7
ABN95134
ID ABN95134 standard; DNA; 1232 BP.
XX
AC ABN95134;
XX
DT 13-AUG-2002 (first entry)
XX
DE Gene #1632 used to diagnose liver cancer.
XX
XX Gene: liver cancer; ds; hepatocellular carcinoma; hepatotropic;
KW metastatic liver tumor; cytostatic; expression profile; disease state;
KW disease progression; drug toxicity; drug efficacy; drug metabolism.
XX
OS Homo sapiens.
XX
PN WO200229103-A2.
XX
PD 11-APR-2002.
XX
XX 02-OCT-2001; 2001WO-US30589.
PF
XX 02-OCT-2000; 2000US-237054P.
PR
XX (GENE-) GENE LOGIC INC.
XX
PI Horne D, Alvares C, Peres-Da-Silva S, Vockley JG;
XX
XX WPI; 2002-426119/45.
XX
PT Diagnosing and detecting the progression of liver cancer.
PT hepatocellular carcinoma or metastatic liver tumor in a patient.
PT Involves detecting the level of expression of two or more genes in a
PT liver tissue sample -
XX
XX Claim 1; SEQ ID NO 1632; 298pp; English.
XX
CC The invention relates to a novel method for diagnosing and detecting the
CC progression of liver cancer, hepatocellular carcinoma or metastatic liver
CC tumor in a patient, and differentiating metastatic liver cancer from
CC hepatocellular carcinoma in a patient, involving detecting the level of
CC expression of two or more genes represented in ABN93503-ABN97455 in a
CC tissue sample. The method of the invention has hepatotropic, and
CC cytostatic activity. The method is useful for diagnosing and detecting
CC the progression of liver cancer, hepatocellular carcinoma and metastatic
CC liver carcinoma in a patient. The method is useful for identifying
CC expression profiles which serve as useful diagnostic markers as well as
CC markers that can be used to monitor disease states, disease progression,
CC drug toxicity, drug efficacy and drug metabolism.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1232 BP; 329 A; 240 C; 251 G; 411 T; 1 other:
Query Match 53.9%; Score 187.4; DB 24; Length 1232;
Best Local Similarity 80.9%; Pred. No. 3.6e-37;
Matches 263; Conservative 1; Mismatches 37; Indels 24; Gaps 3;
OY 4 AACATCACTGCTCTCCAGAGTACCAACATGTCATGTGATGAGCCGCCCTCTGA 63
DB 905 AACATCACTGCTCTCCAGAGTACCAACATGTCATGTGATGAGCCGCCCTCTGA 963
OY 64 CCTCACCCAGCGGACACATGCTTCGGTA-----CCTTTGGGT 102
DB 964 CCTGCGCCAGCGGACACACGCTCTGATGCTGCTTGCCTATGTAGTGTCTTTGGGT 1023
OY 103 CTGTGAGGTCTGTGCA--GGGCTAGTGTCTAACGCCCTTCTGTACACCTTACACG 160
DB 1024 CTGTGAGGTCTGTGCA--GGGCTAGTGTCTAACGCCCTTCTGTACACCTTACACG 1083

QY	161	CAAGAACAACAGTGTGGGCCCTTTGGACACACTAGAACAACTTTTTCATTGACAGTTGC	220
Db	1084	CGAATAATACACGCGTGGGACCCCTTACCTACACAGAAATTTTAAATTGACAGCTTGC	1143
QY	221	AGAAATTGGAGAGTGTGTTTACATTGACATCTTTGGTAAATGACAGTATGTTTGGCA	280
Db	1144	AGAAATTGGAGAGTGTGTTTACATTGACATCTTTGCTTAATGCAATTAGCAATTTATGTTTGGCA	1203
QY	281	TGTATGACTTAATAAATCCTTGAAAT	305
Db	1204	TGTATGACTTAATAAATCCTTGAAAT	1228
RESULT 8			
AAS65569			
ID	AAS65569	standard; cDNA: 1318 BP.	
XX			
AC	AAS65569;		
XX			
DT	13-FEB-2002	(first entry)	
XX			
DE	DNA encoding novel human diagnostic protein #1373.		
XX			
KW	Human; chromosome mapping; gene mapping; gene therapy; forensic;		
XX	food supplement; medical imaging; diagnostic; genetic disorder; ss.		
XX			
OS	Homo sapiens.		
XX			
PN	WO200175067-A2.		
PD			
XX	11-OCT-2001.		
XX			
PE	30-MAR-2001; 2001WO-US08631.		
XX			
PR	31-MAR-2000; 2000US-0540217.		
PR	23-AUG-2000; 2000US-0649167.		
XX			
PA	(HYSE-) HYSEQ INC.		
XX			
PI	Dmanac RT, Liu C, Tang YT;		
XX			
DR	WPI; 2001-639362/73.		
DR	P-PSDB; ABG01382.		
PT	New isolated polynucleotide and encoded polypeptides, useful in		
PT	diagnostics, forensics, gene mapping, identification of mutations		
PT	responsible for genetic disorders or other traits and to assess		
PT	biodiversity -		
XX			
XX			
PS	Claim 1; SEQ ID No 1373; 103pp; English.		
XX			
CC	The invention relates to isolated polynucleotide (I) and		
CC	polypeptide (II) sequences. (I) is useful as hybridisation probes,		
CC	polymerase chain reaction (PCR) primers, oligomers, and for chromosome		
CC	and gene mapping, and in recombinant production of (II). The		
CC	polynucleotides are also used in diagnostics as expressed sequence tags		
CC	for identifying expressed genes. (I) is useful in gene therapy techniques		
CC	to restore normal activity of (II) or to treat disease states involving		
CC	(II). (II) is useful for generating antibodies against it, detecting or		
CC	quantitating a polypeptide in tissue, as molecular weight markers and as		
CC	a food supplement. (II) and its binding partners are useful in medical		
CC	imaging of sites expressing (II). (I) and (II) are useful for treating		
CC	disorders involving aberrant protein expression or biological activity.		
CC	The polypeptide and polynucleotide sequences have applications in		
CC	diagnostics, forensics, gene mapping, identification of mutations		
CC	responsible for genetic disorders or other traits to assess biodiversity		
CC	and to produce other types of data and products dependent on DNA and		
CC	amino acid sequences. AAS64197-AAS94564 represent novel human		
CC	diagnostic coding sequences of the invention.		
CC	Note: The sequence data for this patent did not appear in the printed		
CC	specification, but was obtained in electronic format directly from WIPO		
CC	at ftp.wipo.int/pub/published_pct_sequences.		

XX	Sequence	1318 BP; 346 A; 268 C; 279 G; 424 T; 1 other;
XX	Query Match	53.9%; Score 187.4; DB 23; Length 1318;
XX	Best Local Similarity	80.9%; Pred. No. 3.7e-37;
XX	Matches	263; Conservative 1; Mismatches 37; Indels 24; Gaps 3;
QY	4 AACATCACTGCTCTTCCAGAGTACCAACACTGTCATGTGATGCGCCGCCCTGTAGA	63
DB	991 AACATCACTGCTCTTCCAGAGTACCAACACTGTCATGTGATGCGCCGCCCTGTAGA	1049
QY	64 CCTACCCACGCGAGACATGCTCTCCGGTA-----CCTTGGGT	102
DB	1050 CCTGCGCCACCGGAGACACAGCCCTCTGTAGTCCCTTGGCTATTGATGTTCTTTGGGT	1109
QY	103 CTGTGAGTCTGTCAA--GGCGTAGTGTCAAGCCGCTTGTGCAACCTAATCACTG	160
DB	1110 CTGTGAGTCTGTCAA--GGCGTAGTGTCAAGCCGCTTGTGCAACCTAATCACTG	1169
QY	161 CAAGAACACAGTGTGGGCTTTGGACACATAGAACAACTTTTGAATTGACAGTTGC	220
DB	1170 CGAGAAATACAGCGTGGGACCCCTTACGCCACTACAAAGAAATTTTAAATTTGACAGTTGC	1229
QY	221 AGAATTTGAGAGTGTTTTACATTGATCTTTGCTAAATGACAGTATGATGTTTGA	280
DB	1230 AGAATTTGAGAGTGTTTTACATTGATCTTTGCTAAATGACAGTATGATGTTTGA	1289
QY	281 TGTATGACTTAATAATCCTTGAAAT	305
DB	1290 TGTATGACTTAATAATCCTTGAAAT	1314
RESULT 9	ABX46267/c	
ID	ABX46267	standard; cDNA: 353 BP.
AC	ABX46267;	
XX		
DT	21-FEB-2003	(first entry)
XX		
DE	Bovine EST associated with lactation/muscle/fat deposition #11432.	
XX		
KM	Bovine; ss; EST; expressed sequence tag; lactation; LMPD;	
KM	muscle deposition; fat deposition; genome mapping; gene identification;	
XX	gene analysis; cattle breeding.	
OS	Bos Taurus.	
XX		
PN	US2002137139-A1.	
XX		
PD	26-SEP-2002.	
XX		
PF	24-SEP-2001; 2001US-0960352.	
XX		
PR	12-JAN-1999; 99US-115707P.	
XX	11-JAN-2000; 2000US-0480902.	
PA	(BYAT/) BYATT J C.	
PA	(MATH/) MATHIALAGAN N.	
PA	(TAON/) TAO N.	
XX	(WARR/) WARREN W C.	
XX		
PI	Byatt JC, Mathialagan N, Tao N, Warren WC;	
XX		
DR	WPI: 2003-110599/10.	
XX		
PT	New nucleic acid associated with lactation, and muscle and fat	
PT	deposition, useful for genome mapping, gene identification and	
PT	analysis, cattle breeding, or for genetically improving cattle	
XX		
PS	Claim 2; SEQ ID NO 11432; 245bp; English.	
CC	The invention relates to a purified nucleic acid molecule associated with	

[illegible]

QY	120	GGCGTAGTGC	AAAGCGGTTCTGTACAAACCTAACCTACTGCGAAGAACACAGTGTGGGC	179
Db	286	CTGTATGTGC	TAACGCGTGTCTCTGTACAAACCTAACCTACTGCGAAGAACAGTGTGGGC	345
OY	180	CTTTCGAC	CACTAGACAAACTTTTTCATGTACAGTTGCGAAGATGT -GGAGTGT	238
Db	346	CTTTCACAC	CACTAGACAAACTTTTTCATGTACAGTTGCGAAGATGTGAGAGTGT	405
OY	239	TACATGAT	CTTTTGC 254	
Db	406	TACATGAT	CTTTTGC 421	
RESULT 12				
ID	AAH22395/C			
XX	AAH22395	standard; DNA; 123 BP.		
AC	AAH22395;			
DN	22-AUG-2001	(first entry)		
XX				
DE	Human rac1 genomic fragment probable 3' UTR SEQ ID NO:1.			
XX				
KW	Identification; toxic; hepatotoxic; differential gene expression; NSAID; non-steroidal antiinflammatory drug; ds.			
OS	Homo sapiens.			
XX	WO200138579-A2.			
PN	31-MAY-2001.			
XX				
PE	21-NOV-2000; 2000WO-US32049.			
XX				
PR	22-NOV-1999; 99US-0166923.			
PR	18-FEB-2000; 2000US-0183531.			
XX	20-NOV-2000; 2000US-0717321.			
PA	(CURA-) CURAGEN CORP.			
XX				
PI	Gould-Rothberg BE, DiIppio VA, Ramsehl TM, Gerweln RW;			
DR	WPI; 2001-355948/37.			
XX				
PT	Screening hepatotoxic agent comprises contacting test cell population			
PT	expressing RISKMARKER or INJURYMARKER with agent, comparing expression			
PT	with reference population and identifying difference in expression			
XX	levels -			
PS	Disclosure; Page 7; 76pp; English.			
XX				
CC	The present invention describes a method of screening a test agent for			
CC	hepatotoxicity. The method comprises: (a) providing a test cell			
CC	population comprising a cell capable of expressing one or more nucleic			
CC	acid sequences selected from the group consisting of RISKMARKER 1-8			
CC	and INJURYMARKER 1-10; (b) contacting the test cell population with a			
CC	test agent; (c) measuring expression of one or more of the nucleic			
CC	acid sequences in the test cell population; (d) comparing the			
CC	expression of the nucleic acid sequence in the test cell population to			
CC	the expression of the nucleic acid sequence in a reference cell			
CC	population comprising at least one cell whose exposure status to a			
CC	hepatotoxic agent is known; and (e) identifying a difference in			
CC	expression levels of the RISKMARKER or INJURYMARKER sequences, if			
CC	present, in the test cell population and reference cell population.			
CC	The method is useful for identifying a hepatotoxic agent. The present			
CC	sequence is given in the exemplification of the present invention.			
XX				
SO	Sequence 123 BP; 36 A; 28 C; 30 G; 29 T; 0 other;			
Query Match	35.3%;	Score 123;	DB 22;	Length 123;
Best Local Similarity	100.0%;	Pred. No. 2-2e-21;		
Matches 123;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

Qy	90	GGTACCTTTGGGTCTGTAGGTTCTGTCAAGGCTAGTCTTAACGCCGTTCTGTACAAACC	149
Db	123	GGTACCTTTGGGTCTGTAGGTTCTGTCAAGGCTAGTCTTAACGCCGTTCTGTACAAACC	64
Qy	150	TAACCACTGGGCAACACACAGTGTGGGCGTTTGGACACAGAACAAACCTTTTTCAA	209
Db	63	TAACCACTGGGCAACACACAGTGTGGGCGTTTGGACACAGAACAAACCTTTTTCAA	4
Qy	210	TTG	212
Db	3	TTG	1
RESULT 13			
ID	AA07296		
XX	AA07296 standard; DNA; 3740 BP.		
AC	AA07296;		
DT	21-NOV-2001 (first entry)		
DE	Human reproductive system related antigen DNA Shq ID NO: 9984.		
KX	Human; reproductive system related antigen; reproductive system disorder;		
KM	cancer; gene therapy; ds.		
XX	Homo sapiens.		
PN	W0200155320-A2.		
PD	02-AUG-2001.		
XX			
XX	17-JAN-2001; 2001WO-US01339.		
PR	31-JAN-2000; 2000US-0179065.		
PR	04-FEB-2000; 2000US-0180628.		
PR	24-FEB-2000; 2000US-0184664.		
PR	02-MAR-2000; 2000US-0186350.		
PR	16-MAR-2000; 2000US-0189874.		
PR	17-MAR-2000; 2000US-0190076.		
PR	18-APR-2000; 2000US-0198123.		
PR	19-MAY-2000; 2000US-0205515.		
PR	07-JUN-2000; 2000US-0209467.		
PR	28-JUN-2000; 2000US-0214886.		
PR	30-JUN-2000; 2000US-0215135.		
PR	07-JUL-2000; 2000US-0216647.		
PR	07-JUL-2000; 2000US-0216880.		
PR	11-JUL-2000; 2000US-0217487.		
PR	11-JUL-2000; 2000US-0217496.		
PR	14-JUL-2000; 2000US-0218290.		
PR	26-JUL-2000; 2000US-0220963.		
PR	26-JUL-2000; 2000US-0220964.		
PR	14-AUG-2000; 2000US-0224518.		
PR	14-AUG-2000; 2000US-0224519.		
PR	14-AUG-2000; 2000US-0225266.		
PR	14-AUG-2000; 2000US-0225267.		
PR	14-AUG-2000; 2000US-0225268.		
PR	14-AUG-2000; 2000US-0225270.		
PR	14-AUG-2000; 2000US-0225447.		
PR	14-AUG-2000; 2000US-0225757.		
PR	14-AUG-2000; 2000US-0225758.		
PR	14-AUG-2000; 2000US-0225759.		
PR	18-AUG-2000; 2000US-0226279.		
PR	22-AUG-2000; 2000US-0226681.		
PR	22-AUG-2000; 2000US-0226686.		
PR	22-AUG-2000; 2000US-0227182.		
PR	23-AUG-2000; 2000US-0227009.		
PR	30-AUG-2000; 2000US-0228924.		
PR	01-SEP-2000; 2000US-0228287.		
PR	01-SEP-2000; 2000US-0229343.		
PR	01-SEP-2000; 2000US-0229344.		

PR	01-SEP-2000	2000US-0229345
PR	05-SEP-2000	2000US-0229530
PR	05-SEP-2000	2000US-0229513
PR	06-SEP-2000	2000US-0230437
PR	06-SEP-2000	2000US-0230438
PR	08-SEP-2000	2000US-0231242
PR	08-SEP-2000	2000US-0231243
PR	08-SEP-2000	2000US-0231244
PR	08-SEP-2000	2000US-0231413
PR	08-SEP-2000	2000US-0231414
PR	08-SEP-2000	2000US-0232080
PR	12-SEP-2000	2000US-0232081
PR	12-SEP-2000	2000US-0231968
PR	14-SEP-2000	2000US-0232397
PR	14-SEP-2000	2000US-0232398
PR	14-SEP-2000	2000US-0232399
PR	14-SEP-2000	2000US-0232400
PR	14-SEP-2000	2000US-0232401
PR	14-SEP-2000	2000US-0233063
PR	14-SEP-2000	2000US-0233065
PR	21-SEP-2000	2000US-0234223
PR	21-SEP-2000	2000US-0234274
PR	25-SEP-2000	2000US-0234997
PR	25-SEP-2000	2000US-0234998
PR	26-SEP-2000	2000US-0235484
PR	27-SEP-2000	2000US-0235834
PR	27-SEP-2000	2000US-0235835
PR	29-SEP-2000	2000US-0236527
PR	29-SEP-2000	2000US-0236537
PR	29-SEP-2000	2000US-0236568
PR	29-SEP-2000	2000US-0236569
PR	29-SEP-2000	2000US-0236570
PR	02-OCT-2000	2000US-0237037
PR	02-OCT-2000	2000US-0237038
PR	02-OCT-2000	2000US-0237039
PR	02-OCT-2000	2000US-0237040
PR	13-OCT-2000	2000US-0239393
PR	13-OCT-2000	2000US-0239397
PR	20-OCT-2000	2000US-0240560
PR	20-OCT-2000	2000US-0241121
PR	20-OCT-2000	2000US-0241185
PR	20-OCT-2000	2000US-0241787
PR	20-OCT-2000	2000US-0241808
PR	20-OCT-2000	2000US-0241809
PR	20-OCT-2000	2000US-0244126
PR	01-NOV-2000	2000US-0244617
PR	08-NOV-2000	2000US-0245523
PR	08-NOV-2000	2000US-0245524
PR	08-NOV-2000	2000US-0245525
PR	08-NOV-2000	2000US-0245526
PR	08-NOV-2000	2000US-0246572
PR	08-NOV-2000	2000US-0246578
PR	08-NOV-2000	2000US-0246582
PR	08-NOV-2000	2000US-0246583
PR	08-NOV-2000	2000US-0246609
PR	08-NOV-2000	2000US-0246610
PR	08-NOV-2000	2000US-0246611
PR	11-NOV-2000	2000US-0249613
PR	11-NOV-2000	2000US-0249617
PR	11-NOV-2000	2000US-0249608
PR	11-NOV-2000	2000US-0249609
PR	11-NOV-2000	2000US-0249610
PR	11-NOV-2000	2000US-0249611
PR	11-NOV-2000	2000US-0249612
PR	11-NOV-2000	2000US-0249613
PR	11-NOV-2000	2000US-0249614

PN US2002137139-A1.
 XX
 PD 26-SEP-2002.
 XX
 PF 24-SEP-2001; 2001US-0960352.
 XX
 PR 12-JAN-1999; 99US-115707P.
 PR 11-JAN-2000; 2000US-0480902.
 XX
 PA (BYAT/) BYATT J C.
 PA (MATH/) MATHIALAGAN N.
 PA (TAON/) TAO N.
 PA (WARR/) WARREN W C.
 XX
 PI Byatt JC, Mathialagan N, Tao N, Warren WC;
 DR WPI: 2003-110599/10.
 XX
 XX New nucleic acid associated with lactation, and muscle and fat
 PT deposition, useful for genome mapping, gene identification and
 PT analysis, cattle breeding, or for genetically improving cattle -
 XX
 PS Claim 2; SEQ ID No 4357; 245bp; English.

CC The invention relates to a purified nucleic acid molecule associated with
 CC lactation or muscle and fat deposition (designated LMPD), derived
 CC from cattle, and the LMPD nucleic acid can specifically hybridise to a
 CC second nucleic acid molecule comprising any of 15112 nucleotide
 CC sequences, appearing as ABX34836-ABX49947, or complements of them.
 CC Also included are: (1) a transformed cell having a nucleic acid
 CC comprising an LMPD nucleic acid linked to a promoter and a 3' non-
 CC translated sequence that functions in the cell to cause termination of
 CC transcription and addition of polyadenylated ribonucleotides to a 3' end
 CC of the mRNA molecule; and (2) determining a level or pattern of a
 CC molecule in a bovine cell or tissue comprising: (a) incubating a marker
 CC nucleic acid (comprising any of the 15112 nucleic acid sequences or its
 CC complement or fragment) with a complementary nucleic acid molecule
 CC obtained from the bovine cell or tissue, where hybridisation between the
 CC marker nucleic acid and the complementary nucleic acid permits the
 CC detection of the molecule; and (b) detecting the level or pattern of the
 CC complementary nucleic acid, where the detection of the complementary
 CC nucleic acid is predictive of the level or pattern of the molecule.
 CC The LMPD nucleic acid is used for determining a level or pattern
 CC of a molecule in a bovine cell or tissue. It is useful for genome
 CC mapping, gene identification and analysis, cattle breeding, preparation
 CC of constructs for use in cattle gene expression, or for genetically
 CC improving cattle. The present sequence is one of the 15112 bovine
 CC LMPD EST (expressed sequence tag) nucleic acids.
 CC Note: The present sequence was not shown in the specification but
 CC was obtained in electronic format from the USPTO web site:
 CC seqdata.uspto.gov/sequence.html?DocID=20020137139.

XX
 SQ Sequence 201 BP; 67 A; 34 C; 23 G; 77 T; 0 other;

Query Match 27.4%; Score 95.4; DB 25; Length 201;
 Best Local Similarity 76.5%; Pred. No. 1.9e-14;
 Matches 117; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

QY 174 TTGGGCTTCGACGACATGAGCAACTTTTTCATGTGACAGTGCAGATTTGGAGT 233
 DB 193 TTGAACCTTTCAAAATTAATTAATTTTAAATGACAGTGCAGAAATTTTGAAGT 134
 QY 234 GTTTTACATGATCTTTGCTAATGACATGATGATGTTGATGATGATGATGAT 293
 DB 133 GATTTAGATGATGATTTTGAATGATGATGATGATGATGATGATGATGATGAT 74
 QY 294 AAATCCTTGAATCAATAAAAAAT 326
 DB 73 AAATCTGTGACCCCTCAGATATTAAGATTAAT 41

Search completed: August 25, 2003, 23:47:57
 Job time : 98.0274 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 25, 2003, 19:49:37 ; Search time 723.909 Seconds
(without alignments)
11683.734 Million cell updates/sec

Title: US-09-717-321A-17
Perfect score: 348
Sequence: 1 tgaacacacactgtcttgc.....cttgggaacttgaaaaaaa 348

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 22781392 segs, 12152238056 residues
Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: em_estdb:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estrov:*
6: em_estrpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_hnv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_fod:*
26: em_gss_phg:*
27: em_gss_vr1:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	324	93.1	429	9	A1059212 UI-R-C1-1
C 2	323.4	92.9	493	12	BF554834 UI-R-E0-C
C 3	323.4	92.9	687	12	BM389059 UI-R-D0-C
C 4	321.8	92.5	459	13	BQ780699 UI-R-F0-C

C 5	321.8	92.5	640	13	BQ780117 UI-R-F0-C
C 6	321.4	92.4	477	9	AM520555 UI-R-B0P
C 7	321.4	92.4	489	9	AA859879 UI-R-E0-C
C 8	321.4	92.4	565	12	BI273986 UI-R-CW0-
C 9	321.4	92.4	760	13	BQ191985 UI-R-DN1-
C 10	320.8	92.2	472	13	BQ190077 UI-R-CN1-
C 11	320.8	92.2	615	10	BS380448 UI-R-CT0-
C 12	320.8	92.2	639	10	BE111691 UI-R-BJ1-
C 13	320.8	92.2	705	13	BQ200122 UI-R-EH1-
C 14	318.2	91.4	465	9	AA899757 UI-R-E0-d
C 15	317	91.1	405	10	BE104696 UI-R-BX0-
C 16	317	91.1	412	13	BQ780657 UI-R-FE0-
C 17	317	91.1	517	9	AA859940 UI-R-E0-C
C 18	317	91.1	636	13	BQ780658 UI-R-FE0-
C 19	317	91.1	638	12	BI283790 UI-R-CW0s
C 20	313.8	90.2	469	12	BI276611 UI-R-CX0-
C 21	312.4	89.8	557	12	BS665005 DRABXA06
C 22	310.6	89.3	446	12	BI286743 UI-R-CT0s
C 23	305.4	87.8	463	10	BF548903 UI-R-A0-a
C 24	304.8	87.6	316	9	A1412434 EST240733
C 25	300	86.2	316	9	A1409795 EST238088
C 26	297.2	85.4	407	9	A1009011 EST203462
C 27	297.2	85.4	455	9	A1012196 EST206647
C 28	295.6	84.9	553	9	AA799542 EST189039
C 29	293.6	84.4	427	9	A1178082 EST21741
C 30	292	83.9	431	9	A1410870 EST239163
C 31	285.6	82.1	292	9	A1598992 EST250695
C 32	278	79.9	294	9	A1236740 EST23302
C 33	276.8	79.5	522	14	C06842 C06842 Rat
C 34	273.8	78.7	277	10	BF420446 UI-R-BJ2
C 35	251.2	72.2	277	10	BF420436 UI-R-BJ2-
C 36	246.4	70.8	404	10	BF549054 UI-R-A0-a
C 37	243.4	69.9	644	14	CA339407 NISC_1702
C 38	238.4	68.5	256	9	A1234716 EST21278
C 39	227.6	65.4	394	12	BM022196 I6B0H06.x
C 40	226.6	65.1	322	13	BU757481 UI-1-CFO-
C 41	226.6	65.1	324	13	BU756887 UI-1-CFO-
C 42	226.6	65.1	546	12	BO033391 UI-1-CFO-
C 43	226.6	65.1	555	10	BE996534 UI-M-CG0P
C 44	226	64.9	408	12	BI157798 602923126
C 45	225.8	64.9	404	12	BM502810 1144d06.x

ALIGNMENTS

RESULT 1
LOCUS A1059212/c 429 bp mRNA linear EST 11-FEB-1999
DEFINITION UI-R-C1-1b-e-11-0-UI.s1 UI-R-C1 Rattus norvegicus cDNA clone
UI-R-C1-1b-e-11-0-UI 3', mRNA sequence.

ACCESSION A1059212
VERSION A1059212.1 GI:3332989
KEYWORDS EST.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus

REFERENCE 1 (bases 1 to 429)
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
PubMed 8889548

COMMENT
Contact: Soares, MB
Coordinated Laboratory for Computational Genomics
University of Iowa
375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: bento-soares@iowa.edu

The sequence tag present in the cDNA between the NotI site and the oligo-dT track served to identify it as a clone from the normalized adult lung library. cDNA library Preparation: M. Fatima Bonaldo, Ph.D. Clone distribution: clones will be available through Research Genetics

Seq primer: M13 Forward.

FEATURES
Source Location/Qualifiers

1..429

/organism="Rattus norvegicus"

/mol_type="mRNA"

/strain="Sprague-Dawley"

/db_xref="taxon:10116"

/clone="UI-R-C1-1b-e-11-0-UI"

/dev_stage="Adult"

/lab_host="DH10B (Life Technologies)"

/clone_1lb="UI-R-C1"

/note="Vector: p773D-Pac (Pharmacia) with a modified polylinker. Site.1: Not I; Site.2: Eco RI; The UI-R-C1 library is a subtracted library derived from the UI-R-C0 library, which is a subtracted library derived from the UI-R-A1 and UI-R-E1 libraries. The UI-R-A1 library consisted of a mixture of individually tagged normalized libraries constructed from rat placenta, adult lung, brain, liver, kidney, heart, spleen, ovary, and muscle. The UI-R-E1 library consisted of a mixture of individually tagged normalized libraries constructed from 8, 12 and 18-day embryo. The tag is a string of 3-5 nucleotides present between the Not I site and the oligo-dT track which allows identification of the library of origin of a clone within the mixture. The subtracted library (UI-R-C1) was constructed as follows: PCR amplified cDNA inserts from UI-R-C0 clones from which 3' ESTs had been derived was used as a driver in a hybridization with the UI-R-C0 library in the form of single-stranded circles. The remaining single-stranded circles (subtracted library) was purified by hydroxyapatite column chromatography, converted to double-stranded circles and electroporated into DH10B bacteria (Life Technologies) to generate the UI-R-C1 library. This procedure has been previously described (Bonaldo, Lennon and Soares, Genome Research 6: 791-806, 1996)"

BASE COUNT 130 a 87 c 91 g 121 t
ORIGIN

Query Match 93.18; Score 324; DB 9; Length 429;

Best Local Similarity 95.7%; Pred. No. 4.5e-37;

Matches 333; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

```
1 TGAACATCATCTGCTTGGCAGAGTACCAACACGTGATGATGATGCCGCCCTCT 60
|||||
357 TGAACATCATCTGCTTGGCAGAGTACCAACACGTGATGATGATGCCGCCCTAT 298
|||||
61 AGACCTCACCCAGCGGACACATGCTCCGGTACCTTGGGTCTGTGAGTTCTGTCAG 120
|||||
297 AGACCTCACCCAGCGGACACATGCTCCGGTACCTTGGGTCTGTGAGTTCTGTCAG 238
|||||
121 CGCTAGTGTACAGCGGCTTGTGTACAACTACTACTGCGAAGAACACAGTGTGGCC 180
|||||
237 CGCTAGTGTACAGCGGCTTGTGTACAACTACTACTGCGAAGAACACAGTGTGGCC 178
|||||
181 TTTGACACACTAGAACAACTTTTTCATTTGACAGTTGAGAAATTTGTGAGTTT 240
|||||
177 TTTGACACACTAGAACAACTTTTTCATTTGACAGTTGAGAAATTTGTGAGTTT 118
|||||
241 CATTTGATCTTTTGTCTATGAGTATGAGTATGATGATGATGATGATGATGAT 300
|||||
117 CATTTGATCTTTTGTCTATGAGTATGAGTATGATGATGATGATGATGATGAT 58
|||||
301 TGAATCATTAATAAATAAATAAATGCTTTTGAACCTGAATAAATAA 348
|||||
57 TGAATCATTAATAAATAAATAAATGCTTTTGAACCTGAATAAATAA 10
|||||
```

RESULT 2
BF554834 493 bp mRNA linear EST 12-DEC-2000
LOCUS
DEFINITION
UI-R-E0-cc-c-06-0-UI.r1 UI-R-E0 Rattus norvegicus cDNA clone

ACCESSION
BF554834
VERSION
BF554834.1 GI:11664606
KEYWORDS
EST.
Rattus norvegicus (Norway rat)
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

REFERENCE

1 (bases 1 to 493)
Bonaldo,M.F., Lennon,G. and Soares,M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)

JOURNAL

Medline
PubMed
Contact: Soares, MB
Coordinated Laboratory for Computational Genomics
University of Iowa
375 Newton Road, 4156 MEBR, Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565

Email: Bento-soares@uiowa.edu
cDNA library Preparation: M.B. Soares Lab Clone distribution:
clones will be available through Research Genetics (www.resgen.com)
This clone is also available through the I.M.A.C.E. Consortium at
LNL (info@image.lnl.gov). IMAGE ID-1770064 The following
repetitive elements were found in this cDNA sequence: 463-485,
>AT-richlow-complexity
Seq primer: M13 Forward.

FEATURES
Source Location/Qualifiers

1..493

/organism="Rattus norvegicus"

/mol_type="mRNA"

/strain="Sprague-Dawley"

/db_xref="taxon:10116"

/clone="UI-R-E0-cc-c-06-0-UI"

/dev_stage="embryonic"

/lab_host="DH10B (Life Technologies)"

/clone_1lb="UI-R-E0"

/note="Vector: p773D-Pac (Pharmacia) with a modified polylinker. Site.1: NotI; Site.2: EcoRI; This library consists of a mixture of individually tagged normalized libraries constructed from 8, 12 and 18-day embryo. The tag is a string of 3-5 nucleotides present between the Not I site and the oligo-dT track which allows identification of the library of origin of a clone within the mixture."

BASE COUNT 130 a 109 c 95 g 159 t
ORIGIN

Query Match 92.98; Score 323.4; DB 10; Length 493;

Best Local Similarity 99.7%; Pred. No. 5.1e-37;

Matches 324; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
1 TGAACATCATCTGCTTGGCAGAGTACCAACACGTGATGATGATGCCGCCCTCT 60
|||||
157 TGAACATCATCTGCTTGGCAGAGTACCAACACGTGATGATGATGCCGCCCTCT 216
|||||
61 AGACCTCACCCAGCGGACACATGCTCCGGTACCTTGGGTCTGTGAGTTCTGTCAG 120
|||||
217 AGACCTCACCCAGCGGACACATGCTCCGGTACCTTGGGTCTGTGAGTTCTGTCAG 276
|||||
121 CGCTAGTGTACAGCGGCTTGTGTACAACTACTACTGCGAAGAACACAGTGTGGCC 180
|||||
277 CGCTAGTGTACAGCGGCTTGTGTACAACTACTACTGCGAAGAACACAGTGTGGCC 336
|||||
181 TTTGACACACTAGAACAACTTTTTCATTTGACAGTTGAGAAATTTGTGAGTTT 240
|||||
```

```

Db      337  |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
          TTTGCACCACTAGAACAACTTTTTCATTCAGTTCAGCAAAATGGAGCTGTTTAA 396
Qy      241  CATTGATCTTTTGCATATGACAGTATGTTGCATGATGATTAATAATCCT 300
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      397  CATTATCTTTTGCATATGACAGTTCAGTATGTTTGCATGATGATTAATAATCCT 456
Qy      301  TGAATCATATAAAAAAAAAAAAAA 325
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      457  TGAATCATATAAAAAAAAAAAAAA 481

RESULT 3
BM389059/c 687 bp mRNA linear EST 17-JAN-2002
LOCUS      UI-R-D20-clo-j-07-0-UI.s1 UI-R-D20 Rattus norvegicus cDNA clone
DEFINITION UI-R-D20-clo-j-07-0-UI 3', mRNA sequence.
ACCESSION  BM389059
VERSION     BM389059.1 GI:18189112
KEYWORDS   EST.
SOURCE     Rattus norvegicus (Norway rat)
ORGANISM   Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE  1 (bases 1 to 687)
AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE     Normalization and subtraction: two approaches to facilitate gene
JOURNAL   Genome Res. 6 (9), 791-806 (1996)
MEDLINE   97044477
PUBMED    8889548
COMMENT   Contact: Soares, MB
            Coordinated Laboratory for Computational Genomics
            University of Iowa, 4156 MEBRF, Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: bento-soares@uiowa.edu
            The sequence contained an oligo-dt track that was present in the
            oligonucleotide that was used to prime the synthesis of first
            strand cDNA and therefore this may represent a bonafide poly A
            tail. The sequence tag present in the cDNA between the NotI site
            and the oligo-dt track served to verify it as a clone from the
            non-normalized cartilaginous tumor library cDNA library
            Preparation: M.B. Soares Lab Clone distribution: clones will be
            available through Research Genetics (www.resgen.com)
            Seq primer: M13 Forward
            POLA=yes.

FEATURES
Source      1..687
            Location/Qualifiers
            /organism="Rattus norvegicus"
            /mol_type="mRNA"
            /strain="Sprague-Dawley"
            /db_xref="taxon:10116"
            /clone="UI-R-D20-clo-j-07-0-UI"
            /dev_stage="adult"
            /lab_host="DH10B (Life Technologies)"
            /clone_lib="UI-R-D20"
            /note="Vector: p773D-Pac (Pharmacia) with a modified
            polylinker; Site_1: Not I; Site_2: Eco RI; UI-R-D20 is a
            non-normalized Rat cartilaginous tumor library (RCU)
            constructed in p737 PAC vector according to the procedure
            described by Bonaldo, Lennon & Soares (Genome Research
            genome 6: 791-806, 1996). The oligonucleotide used to
            prime first strand synthesis contained the sequence tag
            CATTCTGTA between the Not I cloning site and dT18
            stretch. The Rat cartilaginous tumor tissue was provided
            by Dr Jeff Stevens at the University of Iowa.
            TAG_LIB=UI-R-D20
            TAG_TISSUE=cartilaginous tumor
            TAG_SEQ=CATTCTGTA"
BASE COUNT 222 a 130 c 163 g 172 t

```

```

ORIGIN
Query Match      92.9%; Score 323.4; DB 12; Length 687;
Best Local Similarity 99.7%; Pred. No. 4,3e-37;
Matches 324; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy      1  TGAACATCAGCTGCTTGGCAGAGTACCAACACTGTCATGTGATGATGCCGCCCTCT 60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      326  TGAACATCAGCTGCTTGGCAGAGTACCAACACTGTCATGTGATGATGCCGCCCTCT 267
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy      61  AGACCTACCCACGCGGACACATGCTTCGCGTACCTTTGGTCTGTGAGTTCTCAAG 120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      266  AGACCTACCCACGCGGACACATGCTTCGCGTACCTTTGGTCTGTGAGTTCTCAAG 207
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy      121  CGCTAGTGTACGCGGCTCTGTACACCTACCTACCTGCGCAAGAACAGTGGGGCC 180
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      206  CGCTAGTGTACGCGGCTCTGTACACCTACCTACCTGCGCAAGAACAGTGGGGCC 147
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy      181  TTTGCACCACTAGAACAACTTTTTCATTCAGTTCAGCAAAATGGAGCTGTTTAA 240
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      146  TTTGCACCACTAGAACAACTTTTTCATTCAGTTCAGCAAAATGGAGCTGTTTAA 87
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy      241  CATTGATCTTTTGCATATGACAGTTCAGTATGTTTGCATGATGATTAATAATCCT 300
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      86  CATTGATCTTTTGCATATGACAGTTCAGTATGTTTGCATGATGATTAATAATCCT 27
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy      301  TGAATCATATAAAAAAAAAAAAAA 325
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      26  TGAATCATATAAAAAAAAAAAAAA 2

RESULT 4
BM780699/c 459 bp mRNA linear EST 26-JUL-2002
LOCUS      UI-R-FF0-cpc-k-22-0-UI.s1 UI-R-FF0 Rattus norvegicus cDNA clone
DEFINITION UI-R-FF0-cpc-k-22-0-UI 3', mRNA sequence.
ACCESSION  BM780699
VERSION     BM780699.1 GI:21989171
KEYWORDS   EST.
SOURCE     Rattus norvegicus (Norway rat)
ORGANISM   Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE  1 (bases 1 to 459)
AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE     Normalization and subtraction: two approaches to facilitate gene
JOURNAL   Genome Res. 6 (9), 791-806 (1996)
MEDLINE   97044477
PUBMED    8889548
COMMENT   Contact: Soares, MB
            Coordinated Laboratory for Computational Genomics
            University of Iowa, 4156 MEBRF, Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: bento-soares@uiowa.edu
            Tissue Procurement: Jeff Stevens
            cDNA library preparation: Dr. M. Bento Soares, University of Iowa
            DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
            Clone Distribution: DISTRIBUTION: Researchers may obtain clones
            from Research Genetics (www.resgen.com).
            Seq primer: M13 FORWARD
            POLA=yes.

FEATURES
Source      1..459
            Location/Qualifiers
            /organism="Rattus norvegicus"
            /mol_type="mRNA"
            /db_xref="taxon:10116"
            /clone="UI-R-FF0-cpc-k-22-0-UI"
            /tissue_type="Mixed tissues"

```

```
/dev_stage="Adult"
/lab_host="DH10B (Life Technologies) (T1 phage resistant)"
/clone_lib="UI-R-FF0"
/notes="Vector: p7713-Pac (Pharmacia) with a modified
polylinker. Site.1: EcoR I; Site.2: Not I; UI-R-FF0 is a
subtracted cDNA library containing the following tissue(s)
): Normal cartilage and SR-JWS Tumor line. The
subtraction was made according to Bonaldo, Lennon and
Soares, Genome Research, 6:791-806, 1996. The
oligonucleotide used to prime the synthesis of
first-strand cDNA contains a library tag sequence that is
located between the Not I site and the (dr)18 tail. The
sequence tags for these libraries are: CTATGGAGCG,
CATTCCTGTA.
TAG_LIB=UI-R-FF0
TAG_TISSUE=cartilage
TAG_SBO=CTATGGAGCG"

BASE COUNT      145 a      91 c      103 g      120 t

Query Match      92.5%: Score 321.8; DB 13; Length 459;
Best Local Similarity 99.4%: Pred. NO. 9e-37;
Matches 323; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

1 TGAACATCAGTCTGTGCGAGATACCAACATGTCATGATGATGCGCCCTCT 60
|||||
325 TGAACATCAGTCTGTGCGAGATACCAACATGTCATGATGATGCGCCCTCT 266
61 AGACCTCACCCAGCGGACACATGCTCCGTAACCTTGGGTCTGTGAGGTTCTGTCAG 120
|||||
265 AGACCTCACCCAGCGGACACATGCTCCGTAACCTTGGGTCTGTGAGGTTCTGTCAG 206
121 CGCTAGTGTAAAGCGCTTGTACAACTTACTGACATGCGCAAGAACAGTGTGGGCC 180
|||||
205 CGCTAGTGTAAAGCGCTTGTACAACTTACTGACATGCGCAAGAACAGTGTGGGCC 146
181 TTTGACCACTAGAACAACTTTTTCATTTGACAGTTGAGAAATTTGGAGTGTTTTA 240
|||||
145 TTTGACCACTAGAACAACTTTTTCATTTGACAGTTGAGAAATTTGGAGTGTTTTA 86
241 CATTTGATCTTTTCTAATGACAGTTAGCATGATGATGATGATGATGATGATGAT 300
|||||
85 CATTTGATCTTTTCTAATGACAGTTAGCATGATGATGATGATGATGATGATGAT 26
301 TGAATCATTAATAAAAAAAAAA 325
|||||
25 TGAATCATTAATAAAAAAAAAA 1

RESULT 5
B0780117/c      640 bp      mRNA      linear      EST 26-JUL-2002
LOCUS DEFINITION
UI-R-FF0-cow-j-10-0-UI.s1 UI-R-FF0 Rattus norvegicus cDNA clone
B0780117
UI-R-FF0-cow-j-10-0-UI 3', mRNA sequence.
ACCESSION      B0780117.1 GI:21988589
VERSION
KEYWORDS
SOURCE
ORGANISM
Rattus norvegicus (Norway rat)
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 640)
Bonaldo, M.F., Lennon, G. and Soares, M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)
JOURNAL MEDLINE 97044477
PUBMED 8889548
COMMENT
Contact: Soares, MB
Coordinated Laboratory for Computational Genomics
University of Iowa
375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA
```

```
Tel: 319 335 8250
Fax: 319 335 9565
Email: bento-soares@uiowa.edu
Tissue Procurement: Jeff Stevens
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
DNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: DISTRIBUTION: Researchers may obtain clones
from Research Genetics (www.resgen.com).
Seq primer: M13 FORWARD
POLYA=Yes.

FEATURES
source
location/Qualifiers
1..640
/organism="Rattus norvegicus"
/mol_type="mRNA"
/db_xref="taxon:10116"
/clone="UI-R-FF0-cow-j-10-0-UI"
/tissue_type="Mixed tissues"
/dev_stage="Adult"
/lab_host="DH10B (Life Technologies) (T1 phage resistant)"
/clone_lib="UI-R-FF0"
/notes="Vector: p7713-Pac (Pharmacia) with a modified
polylinker. Site.1: EcoR I; Site.2: Not I; UI-R-FF0 is a
subtracted cDNA library containing the following tissue(s)
): Normal cartilage and SR-JWS Tumor line. The
subtraction was made according to Bonaldo, Lennon and
Soares, Genome Research, 6:791-806, 1996. The
oligonucleotide used to prime the synthesis of
first-strand cDNA contains a library tag sequence that is
located between the Not I site and the (dr)18 tail. The
sequence tags for these libraries are: CTATGGAGCG,
CATTCCTGTA.
TAG_LIB=UI-R-FF0
TAG_TISSUE=rat SRC-JWST tumor line
TAG_SBO=CATTCCTGTA"

BASE COUNT      208 a      121 c      152 g      159 t

Query Match      92.5%: Score 321.8; DB 13; Length 640;
Best Local Similarity 99.4%: Pred. NO. 7.5e-37;
Matches 323; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

1 TGAACATCAGTCTGTGCGAGATACCAACATGTCATGATGATGCGCCCTCT 60
|||||
326 TGAACATCAGTCTGTGCGAGATACCAACATGTCATGATGATGCGCCCTCT 267
61 AGACCTCACCCAGCGGACACATGCTCCGTAACCTTGGGTCTGTGAGGTTCTGTCAG 120
|||||
266 AGACCTCACCCAGCGGACACATGCTCCGTAACCTTGGGTCTGTGAGGTTCTGTCAG 207
121 CGCTAGTGTAAAGCGCTTGTACAACTTACTGACATGCGCAAGAACAGTGTGGGCC 180
|||||
206 CGCTAGTGTAAAGCGCTTGTACAACTTACTGACATGCGCAAGAACAGTGTGGGCC 147
181 TTTGACCACTAGAACAACTTTTTCATTTGACAGTTGAGAAATTTGGAGTGTTTTA 240
|||||
146 TTTGACCACTAGAACAACTTTTTCATTTGACAGTTGAGAAATTTGGAGTGTTTTA 87
241 CATTTGATCTTTTCTAATGACAGTTAGCATGATGATGATGATGATGATGATGAT 300
|||||
86 CATTTGATCTTTTCTAATGACAGTTAGCATGATGATGATGATGATGATGATGAT 27
301 TGAATCATTAATAAAAAAAAAA 325
|||||
26 TGAATCATTAATAAAAAAAAAA 2

RESULT 6
AM520555/c      477 bp      mRNA      linear      EST 06-MAR-2000
LOCUS DEFINITION
UI-R-BJ0p-afx-g-09-0-UI.s1 UI-R-BJ0p Rattus norvegicus cDNA clone
UI-R-BJ0p-afx-g-09-0-UI 3', mRNA sequence.
ACCESSION      AM520555
```

FEATURES	SOURCE	Location/Qualifiers
1.	477	
		/organism="Rattus norvegicus"
		/mol_type="mRNA"
		/strain="Sprague-Dawley"
		/db_xref="taxon:10116"
		/clone="UI-R-BJ0P-afix-g-09-0-UI"
		/dev_stage="adult"
		/lab_host="DH10B (Life Technologies)"
		/clone_lib="UI-R-BJ0P"
		/note="Vector: pUT73b-Pac (pharmacia) with a modified polylinker. Site.1: Not I; Site.2: Eco RI. The UI-R-BJ0P library is a subcloned library derived from the UI-R-Abl UI-R-Abl, UI-R-AC1, UI-R-AD1, UI-R-AE1, UI-R-AF1, and UI-R-AG1 libraries. These libraries represent tissues from rat atrium at 16.5 dpc, ventricle at 16.5 dpc, AV canal at 16.5 dpc, atrium at 15 dpc, ventricle at 15 dpc, AV canal at 15 dpc, and ventricle at 13 dpc. The tag is a string of 5-6 nucleotides present between the Not I site and the oligo-dT track. The library was constructed as described by Bonaldo, Lennon and Soares, Genome Research 6: 791-806, 1996.
		TAG_Lib-UI-R-BJ0P
		TAG_TISSUE-AV canal at 15 dpc
		TAG_SEQ-GAAGG"
BASE COUNT	157 a	93 c 103 g 124 t
ORIGIN		
Query Match	92.4%	Score 321.4; DB 9; Length 477;
Best Local Similarity	99.7%	Pred. No. 1e-36;
Matches 322; Conservative	0; Mismatches	1; Indels 0; Gaps 0
QY	1	TGAAACATCATCTGTTTCCAGAGTACCAACACTGTCATGATGATGATCGCCGCCCTCT 60
Db	323	TGAAACATCATCTGTTTCCAGAGTACCAACACTGTCATGATGATGATCGCCGCCCTCT 264
QY	61	AGACCTCACCCACGCGGACACATGCTTCCGATACCTTTGGTCTGTGAGTGTCTGTCAAG 120
Db	263	AGACCTCACCCACGCGGACACATGCTTCCGATACCTTTGGTCTGTGAGTGTCTGTCAAG 204
QY	121	CGCTAGTCTTACGCCGTTTCTGTATACACTTAATCTACTGGACAGAACACATGTTGGCC 180

RESULT 7	
AA859879/c	
LOCUS	AA859879 489 bp mRNA linear EST 03-JUL-1998
DEFINITION	UI-R-E0-cc-c-06-0-UI.s1 UI-R-E0 Rattus norvegicus cDNA clone
UI-R-E0-cc-c-06-0-UI 3'	similar to dbj D25274 HUMPO2ST9 Human
randomly sequenced mRNA,	EST.
AA859879	AA859879.1 GI:4230416
ACCESSION	
VERSION	
KEYWORDS	
SOURCE	Rattus norvegicus (Norway rat)
ORGANISM	Rattus norvegicus
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
	Mammalia; Eutheria; Rodentia; Rodentia; Sciurognathi; Muridae; Murinae;
	Rattus.
REFERENCE	1 (bases 1 to 489)
AUTHORS	Bonaldi,M.F., Lennon,G. and Soares,M.B.
TITLE	Normalization and subtraction: two approaches to facilitate gene
	discovery
JOURNAL	Genome Res. 6 (9), 791-806 (1996)
MEDLINE	97044477
PUBMED	8889548
COMMENT	On Mar 10, 1998 this sequence version replaced q1:2949399.
	8889548

```

FEATURES
source
location/Qualifiers
1. 489
/mol_isism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-E0-cc-c-06-0-UI"
/day_stage="embryonic"
/lab_host="DH10B (Life Technologies)"
/clone_lib="UI-R-E0"
/notes="Vector: p7773D-Pac (Pharmacia) with a modified
polylinker. Site.1: NotI; Site.2: EcoRI. This library
consists of a mixture of individually tagged normalized
libraries constructed from 8, 12 and 18-day embryo. The
tag is a string of 3-5 nucleotides present between the
Not I site and the oligo-dr track which allows
identification of the library of origin of a clone within
the mixture."
BASE COUNT
158 a 96 c 107 g 128 t
ORIGIN

```

```
Query Match          92.4%; Score 321.4; DB 9; Length 489;
Best Local Similarity 99.7%; Pred. No. 9.9e-37;
Matches 322; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 TGAACATCAGTCGTCTGGCAGAGTACCAACACTGTCATGTGATGCGCCGCCCTCT 60
    |||
DB 323 TGAACATCAGTCGTCTGGCAGAGTACCAACACTGTCATGTGATGCGCCGCCCTCT 264
    |||
OY 61 AGACCTCACCCACGCGGACACATGCTCCGATCCTTGGGTCTGTGAGGTTCTGTCAG 120
    |||
DB 263 AGACCTCACCCACGCGGACACATGCTCCGATCCTTGGGTCTGTGAGGTTCTGTCAG 204
    |||
OY 121 CGGTAGTGTCTAACGCCGTTCTGTACAACTCACTCACTGGCAAGAACAGTGTGGGCC 180
    |||
DB 203 CGGTAGTGTCTAACGCCGTTCTGTACAACTCACTCACTGGCAAGAACAGTGTGGGCC 144
    |||
OY 181 TTTTCGACCACTAGAACAACTTTTTCATTTGACAGTTCGAGATTTGGAGTTCTTTA 240
    |||
DB 143 TTTTCGACCACTAGAACAACTTTTTCATTTGACAGTTCGAGATTTGGAGTTCTTTA 84
    |||
OY 241 CATTGATCTTTTGTCTAATGACAGTATGTTTTCATGATGATTAATTAATCTT 300
    |||
DB 83 CATTGATCTTTTGTCTAATGACAGTATGTTTTCATGATGATTAATTAATCTT 24
    |||
OY 301 TGAATCATTAATAAAAAAAAAA 323
    |||
DB 23 TGAATCATTAATAAAAAAAAAA 1
    |||

RESULT 8
BI273986/c 565 bp mRNA linear EST 18-JUL-2001
LOCUS BI273986
DEFINITION UI-R-CWO-bwb-c-12-0-UI.s1 UI-R-CWO Rattus norvegicus cDNA clone
            UI-R-CWO-bwb-c-12-0-UI 3', mRNA sequence.
ACCESSION BI273986
VERSION BI273986.1 GI:14884485
KEYWORDS EST.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Cranialta; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE 1 (bases 1 to 565)
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
PUBMED 8889548
COMMENT Contact: Soares, MB
            Coordinated Laboratory for Computational Genomics
            University of Iowa
            375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: bento-soares@uiowa.edu
            The sequence contained an oligo-dT track that was present in the
            oligonucleotide that was used to prime the synthesis of first
            strand cDNA and therefore this may represent a bonafide poly A
            tail. The sequence tag present in the cDNA between the NotI site
            and the oligo-dT track served to identify it as a clone from the
            normalized rat aorta pool library cDNA library preparation: M.B.
            Soares lab clone distribution: clones will be available through
            Research Genetics (www.resgen.com)
            Seq primer: M13 forward
            POLYA=yes.
            Location/Qualifiers
            1..565
            /organism="Rattus norvegicus"
            /mol_type="mRNA"
            /strain="Sprague-Dawley"
            /db_xref="taxon:10116"
```

```
/clone="UI-R-CWO-bwb-c-12-0-UI"
/dev_stage="ADULT"
/lab_host="DH10B (Life Technologies)"
/clone_1lb="UI-R-CWO"
/clone_2lb="UI-R-CWO"
/notes="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site.1: Not I; Site.2: Eco RI; The UI-R-CWO
library is a normalized library constructed from the
following rat aorta tissues: embryonic day 19, embryonic
day 21, adult day 1, adult day 12, adult day 75, adult day
200. For a detailed description of the library from which
this clone was derived, please visit our web site at
rategen.eng.uiowa.edu. The subtraction has been previously
described in (Bonaldo, Lennon, and Soares, Genome Research
6:791-806, 1996)
TAG_LIB=UI-R-CWO
TAG_TISSUE=rat aorta pool
TAG_SEQ=CTGTAGATC"

BASE COUNT 188 a 107 c 123 g 146 t 1 others
ORIGIN

Query Match          92.4%; Score 321.4; DB 12; Length 565;
Best Local Similarity 99.4%; Pred. No. 9.1e-37;
Matches 322; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 TGAACATCAGTCGTCTGGCAGAGTACCAACACTGTCATGTGATGCGCCGCCCTCT 60
    |||
DB 324 TGAACATCAGTCGTCTGGCAGAGTACCAACACTGTCATGTGATGCGCCGCCCTCT 265
    |||
OY 61 AGACCTCACCCACGCGGACACATGCTCCGATCCTTGGGTCTGTGAGGTTCTGTCAG 120
    |||
DB 264 AGACCTCACCCACGCGGACACATGCTCCGATCCTTGGGTCTGTGAGGTTCTGTCAG 205
    |||
OY 121 CGGTAGTGTCTAACGCCGTTCTGTACAACTCACTCACTGGCAAGAACAGTGTGGGCC 180
    |||
DB 204 CGGTAGTGTCTAACGCCGTTCTGTACAACTCACTCACTGGCAAGAACAGTGTGGGCC 145
    |||
OY 181 TTTTCGACCACTAGAACAACTTTTTCATTTGACAGTTCGAGATTTGGAGTTCTTTA 240
    |||
DB 144 TTTTCGACCACTAGAACAACTTTTTCATTTGACAGTTCGAGATTTGGAGTTCTTTA 85
    |||
OY 241 CATTGATCTTTTGTCTAATGACAGTATGTTTTCATGATGATTAATTAATCTT 300
    |||
DB 84 CATTGATCTTTTGTCTAATGACAGTATGTTTTCATGATGATTAATTAATCTT 25
    |||
OY 301 TGAATCATTAATAAAAAAAAAA 324
    |||
DB 24 TGAATCATTAATAAAAAAAAAA 1
    |||

RESULT 9
B0191985/c 760 bp mRNA linear EST 30-APR-2002
LOCUS B0191985
DEFINITION UI-R-DRI-cky-h-11-0-UI.s1 UI-R-DRI Rattus norvegicus cDNA clone
            UI-R-DRI-cky-h-11-0-UI 3', mRNA sequence.
ACCESSION B0191985
VERSION B0191985.1 GI:20367536
KEYWORDS EST.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Cranialta; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE 1 (bases 1 to 760)
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
PUBMED 8889548
COMMENT Contact: Soares, MB
            Coordinated Laboratory for Computational Genomics
            University of Iowa
            375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA
```


Tel: 319 335 8250
Fax: 319 335 9565

Email: bento-soares@uiowa.edu

The sequence contained an oligo-dT track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. The sequence tag present in the cDNA between the NotI site and the oligo-dT track served to verify it as a clone from the normalized osteoblast library cDNA library preparation: M.B. Soares Lab Clone distribution: clones will be available through Research Genetics (www.resgen.com)
Seq primer: M13 Forward
POLYA=yes.

FEATURES

source
Location/Qualifiers
1..760
/organism="Rattus norvegicus"
/mol_type="mRNA"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-DRI-cky-h-11-0-UI"
/dev_stage="adult"
/lab_host="DH10B (Life Technologies)"
/clone_lib="UI-R-DRI"
/note="Vector: pRT3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-DRI library is a normalized Rat Osteoblast library (nREO) constructed in pRT37 vector according to the procedure described by Bonaldo, Lennon & Soares (Normalization and Subtraction: Two Approaches to Facilitate Gene Discovery. Genome Research 6: 791-806, 1996). The oligonucleotide used to prime first strand synthesis contained the sequence tag AGATATCA between the Not I cloning site and dT18 stretch. The Rat Osteoblast tissue was provided by Lian & Stein of the University of Massachusetts Medical School
TAG_LIB=UI-R-DRI
TAG_TISSUE=osteoblast
TAG_SEQ=AGATATCA"
BASE COUNT 244 a 143 c 173 g 198 t 2 others
ORIGIN

Query Match 92.4%; Score 321.4; DB 13; Length 760;

Best Local Similarity 99.4%; Pred. No. 7.7e-37;

Matches 322; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

QY 1 TGAACATCATCTGCTTGGACAGTACCAACATGTCATTTGATGAGCGCCCTCT 60
   |||||||
DB 324 TGAACATCATCTGCTTGGACAGTACCAACATGTCATTTGATGAGCGCCCTCT 265
   |||||||
QY 61 AGACCTCAACCCAGCGGACACATGCTCCGCTACCTTTGGGTCTGTGAGTCTGTCAAG 120
   |||||||
DB 264 AGACCTCAACCCAGCGGACACATGCTCCGCTACCTTTGGGTCTGTGAGTCTGTCAAG 205
   |||||||
QY 121 CGCTAGTGTAAAGCCGCTTGTCTACACCTAACTACAGTGAAGAACACAGTGTGGGCC 180
   |||||||
DB 204 CGCTAGTGTAAAGCCGCTTGTCTACACCTAACTACAGTGAAGAACACAGTGTGGGCC 145
   |||||||
QY 181 TTTGACACATGAGCAAACTTTTTCATTTGACATGACAGTTGAGAGTGTGTTTAA 240
   |||||||
DB 144 TTTGACACATGAGCAAACTTTTTCATTTGACATGACAGTTGAGAGTGTGTTTAA 85
   |||||||
QY 241 CATTTGATCTTTTGTCTAATGACATGATGATGTTTGCATGATGACTTAATATCT 300
   |||||||
DB 84 CATTTGATCTTTTGTCTAATGACATGATGATGTTTGCATGATGACTTAATATCT 25
   |||||||
QY 301 TGAATCATTAATAAAAAAAAAA 324
   |||||||
DB 24 TGAATCATTAATAAAAAAAAAA 1

```

RESULT 10
LOCUS B0190077c 472 bp mRNA linear EST 30-APR-2002

DEFINITION UI-R-CNI-cjs-h-07-0-UI.s3 UI-R-CNI Rattus norvegicus cDNA clone
ACCESSION UI-R-CNI-cjs-h-07-0-UI 3', mRNA sequence.
VERSION B0190077
KEYWORDS B0190077.1 GI:20365628
SOURCE EST.
ORGANISM Rattus norvegicus (Norway rat)
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE 1 (bases 1 to 472)
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
PubMed 8889548

COMMENT
Contact: Soares, MB
Coordinated Laboratory for Computational Genomics
University of Iowa
375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: bento-soares@uiowa.edu
The sequence contained an oligo-dT track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. The sequence tag present in the cDNA between the NotI site and the oligo-dT track served to identify it as a clone from the normalized rat eye library cDNA library preparation: M.B. Soares Lab Clone distribution: clones will be available through Research Genetics (www.resgen.com)
Seq primer: M13 Forward
POLYA=yes.

FEATURES

source

Location/Qualifiers
1..472
/organism="Rattus norvegicus"
/mol_type="mRNA"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-CNI-cjs-h-07-0-UI"
/dev_stage="adult"
/lab_host="DH10B (Life Technologies)"
/clone_lib="UI-R-CNI"
/note="Vector: pRT3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-CNI library is a subtracted library derived from the following pool of seven normalized rat libraries: normalized rat seminal vesicles, normalized rat penis, normalized rat bladder, normalized rat cervix, normalized rat brown adipose, normalized rat fundus, and normalized rat salivary gland. It was constructed according to the procedure described by Bonaldo, Lennon & Soares (Genome Research 6: 791-806, 1996). For construction of the CNI library, plasmid DNA from the pool of normalized libraries was electroporated into competent bacteria for the production of single-stranded circular DNA. This was then used as a tracer in a subtractive hybridization with a driver (PCR amplified inserts from a plasmid DNA template preparation) comprising: a) a pool of about 34,000 clones from the Rat Unigene Set corresponding to plates R-5-AA-NN excluding plates R-5-MM and MN. This pool represented 40% of the final driver population. b) a pool of about 29,000 clones from subtracted libraries C40 and C41 corresponding to plates R-CA0-ANV through R-CA0-AXS, R-CA0-AXZ through R-CA0-BAZ, R-CA0-BFE through R-CA0-BHY, R-CA0-BUS, R-CA0-BKE, R-CA0-BKG-H, R-CA0-BKJ-R, R-CA0-BKP through R-CA0-BKS, R-CA0-BKU-V, R-CA0-BLX through R-CA0-BMA, R-CA0-BMC through R-CA0-BME, R-CA0-BNS, R-CA0-BOB through R-CA0-BQJ, R-CA0-BPA through R-CA0-BPG, R-CA1-BBA through R-CA1-BDA, R-CA1-BHZ through R-CA1-BUF, R-CA1-BUR, R-CA1-BUT through R-CA1-BKB, R-CA1-BKD, R-CA1-BKF, R-CA1-BKI, R-CA1-BKT, R-CA1-BLF, R-CA1-BLH through

R-CA1-BLN, R-CA1-BIS, R-CA1-BLU-V, R-CA1-BNR, and R-CA1-BLE. The resulting pool represented 20% of the final driver population. c) a pool of about 15,000 clones from non-normalized libraries CS0s, CT0s, CU0s, CM0s, CX0s and normalized libraries CS0, CT0, CU0, CM0, and CX0 corresponding to plates R-CS0s-CBD through R-CS0s-CBO, R-CT0s-CAM through R-CT0s-CAX, R-CU0s-CBP through R-CU0s-CCA, R-CM0s-CCB through R-CM0s-CCM, R-CX0s-CCN through R-CX0s-CCX, R-CS0-BSD, R-CT0-BID through R-CT0-BTV, R-CS0-BVM, R-CT0-BTW through R-CT0-BUD, R-CT0-BVN, R-CU0-BUQ through R-CU0-BVL, R-CM0-BVJ through R-CM0-BMP, R-CM0-BXN through R-CM0-BXO, R-CX0-BMQ through R-CX0-BXM. The resulting pool represented 5% of the final driver population. d) a pool of about 5,000 clones (1,000 from non-normalized eye library CV0 and 4,000 from normalized eye library CV1) corresponding to plates R-CV0-BRH through R-CV0-BRR, R-CV1-BRS through R-CV1-BSC, R-CV1-BSE through R-CV1-BTC, and R-CV1-BVO through R-CV1-BVU. This pool represented about 5% of the final driver population. e) A pool of about 10,000 clones from subtracted library BS2, BV0 and BV0P (7-9.5 kb cDNA library fraction from rat whole embryo), and BX0 (0.5-7kb cDNA library fraction from rat whole embryo) corresponding to plates R-BS2-BDB through R-BS2-BFB, R-BV0-ANR through R-BV0-ANR, R-BV0P-AOI through R-BV0P-AOX, and R-BX0-AOY through R-BX0-ASH. The resulting pool represented 5% of the final driver population. f) a pool of about 7,000 clones from the seven non-normalized libraries that make up the tracer including CV0, CZ0, DA0, DB0, DCO, DD0, and DE0 corresponding to plates R-CY0-BXP through R-CY0-BXZ, R-CZ0-BYA through R-CZ0-BYI, R-CZ0-BAB-C, R-DA0-BXJ through R-DA0-BYP, R-DA0-BZD through R-DA0-BZH, R-DB0-BVQ through R-DB0-BZA, R-DC0-BZI through R-DC0-CAZ through R-DC0-CBN, R-DD0-BER through R-DD0-CAZ, R-DD0-CBB-C, and R-DE0-CAB through R-DE0-CAL. The resulting pool represented about 10% of the final driver population. g) a pool of about 2,000 clones from the pool of normalized libraries, CN0, that makes up the tracer. The corresponding plates are R-CN0-BKR through R-CN0-BLD, R-CN0-BIG, R-CN0-BLP through R-CN0-BLT, R-CN0-BLM-X, R-CN0-BMB, and R-CN0-BMP through R-CN0-BML. This pool represented 5% of the final driver population. h) a pool of the 28 most abundant clones in the CN0 pool corresponding to the following addresses: bkw-a-09-0-UI, bkw-b-09-0-UI, bkw-b-11-0-UI, bkw-b-10-0-UI, bkw-d-01-0-UI, bkw-d-06-0-UI, bkw-g-08-0-UI, bkw-h-12-0-UI, bky-a-05-0-UI, bky-a-06-0-UI, bky-a-11-0-UI, bky-c-06-0-UI, bky-c-09-0-UI, bky-d-10-0-UI, bla-a-01-0-UI, bla-a-02-0-UI, bla-f-04-0-UI, bla-g-07-0-UI, bla-g-12-0-UI, blb-a-12-0-UI, blb-f-02-0-UI, blc-a-11-0-UI, blc-e-95-0-UI, bld-1-08-0-UI, bld-f-02-0-UI, blq-h-04-0-UI, blr-a-05-0-UI, blt-f-08-0-UI. This pool represented 5% of the final driver population. i) One abundant CN0 clone (corresponding to the address bky-a-11-0-UI) was digested with Not I and Eco RI and the resulting insert was gel purified. This purified insert was added directly to the driver so that it represented 5% of the final driver population.

TAG_LIB-UI-R-CNI
TAG_TISSUE-rat eye
TAG_SEQ-CAGCC"

BASE COUNT 154 a 93 c 102 g 123 t

ORIGIN

Query Match 92.28; Score 320.8; DB 13; Length 472;
Best Local Similarity 99.48; Pired. No. 1.2e-36;
Matches 322; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGAACATCTACTGCTTGCAGACTACCAACACGTATGTATGATGCCGCCCTCT 60
|||||
DB 324 TGAACATCTACTGCTTGCAGACTACCAACACGTATGTATGATGCCGCCCTCT 265
|||||

QY 61 AGACCTCACCACGACACATGCTTCGGTACCTTGGGTCTGTGAGTTCGTCAAG 120

DB 264 AGACCTCACCACGACACATGCTTCGGTACCTTGGGTCTGTGAGTTCGTCAAG 205
|||||

QY 121 CGGTAGTGTCTACGCGCTTGTGTACAACCTTACTGTGCAAGAACACAGTTGGGCC 180
|||||

DB 204 CGGTAGTGTCTACGCGCTTGTGTACAACCTTACTGTGCAAGAACACAGTTGGGCC 145
|||||

QY 181 TTTCGACCACTAGAACAACTTTTTCATTTGACAGTTGCGAATTTGGAGTTTCTTA 240
|||||

DB 144 TTTCGACCACTAGAACAACTTTTTCATTTGACAGTTGCGAATTTGGAGTTTCTTA 85
|||||

QY 241 CATTGATCTTCTGCTAATGAGTAGAGTAGATGTTTTCATTTGACAGTTTTCATTTTCCT 300
|||||

DB 84 CATTGATCTTCTGCTAATGAGTAGAGTAGATGTTTTCATTTGACAGTTTTCATTTTCCT 25
|||||

QY 301 TGAATCATTAATAAAAAAAAAAAAAA 324
|||||

DB 24 TGAATCATTAATAAAAAAAAAAAAAA 1
|||||

RESULT 11
BG380448/c 615 bp mRNA linear EST 12-MAR-2001
LOCUS
DEFINITION
UI-R-CT0-btx-d-05-0-UI.s1 UI-R-CT0 Rattus norvegicus cDNA clone
BG380448
VERSION
BG380448.1 GI:13304920
KEYWORDS
EST.
Rattus norvegicus (Norway rat)
Rattus norvegicus
Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
COMMENT

1 (bases 1 to 615)
Bonaldo,M.F., Lennon,G. and Soares,M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)
8889348
Contact: Soares, MB
Coordinated Laboratory for Computational Genomics
University of Iowa
375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9365
Email: bento-soares@uiowa.edu

The sequence contained an oligo-dT track that was present in the
oligonucleotide that was used to prime the synthesis of first
strand cDNA and therefore this may represent a bonafide poly A
tail. The sequence tag present in the cDNA between the NotI site
and the oligo-dT track served to verify it as a clone from the
normalized rat brain pool library cDNA library Preparation: M.B.
Soares Lab Clone distribution: clones will be available through
Research Genetics (www.resgen.com)
Seq primer: M13 Forward
POLYA-yes.

FEATURES
SOURCE

Location/Qualifiers
1. 615
/organism="Rattus norvegicus"
/mol_type="mRNA"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-CT0-btx-d-05-0-UI"
/dev_stage="ADULT"
/lab_host="DH10B (Life Technologies)"
/clone_idb="UI-R-CT0"
/note="Vector: p773D-Pac (Pharmacia) with a modified
polylinker: Site_1: Not I; Site_2: Eco RI; The UI-R-CT0
library is a normalized library constructed from the
following rat brain tissues: embryonic day 17, embryonic
day 19, embryonic day 21, adult day 1, adult day 12, adult
day 75, adult day 200. For a detailed description of the

TITLE Normalization and subtraction: two approaches to facilitate gene discovery

JOURNAL Genome Res. 6 (9), 791-806 (1996)

MEDLINE 97044477

PUBMED 8889548

COMMENT Contact: Soares, MB
Coordinated Laboratory for Computational Genomics
University of Iowa
375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: bento-soares@uiowa.edu

The sequence contained an oligo-dT track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a Bonafide poly A tail. The sequence tag present in the cDNA between the NotI site and the oligo-dT track served to verify it as a clone from the normalized duodenum library cDNA library Preparation: M.B. Soares Lab Clone distribution: clones will be available through Research Genetics (www.resgen.com)
Seq primer: M13 Forward
POLYA=Yes.

FEATURES

source

1..705

location/Qualifiers

/organism="Rattus norvegicus"

/mol_type="mRNA"

/strain="Sprague-Dawley"

/db_xref="taxon:10116"

/clone="UI-R-EB1-clf-1-07-0-UI"

/dev_stage="adult"

/lab_host="DH10B (Life Technologies)"

/clone_lib="UI-R-EB1"

/note="Vector: p773D-Pac (Pharmacia) with a modified polylinker: Site_1: Not I; Site_2: Eco RI; UI-R-EB1 is a non-normalized Rat Duodenum library (RDV) constructed in p7737 PAC vector according to the procedure described by Bonaldo, Lennon & Soares (Genome Research Genome 6: 791-806, 1996). The oligonucleotide used to prime first strand synthesis contained the sequence tag TGTGGTTCAT between the Not I cloning site and dT18 stretch. The Rat Duodenum tissue was provided by Tom Freeman of the Sanger Center.

TAG LIB-UI-R-EB1

TAG TISSUE=duodenum

TAG_SEQ=TGTGGTTCAT"

BASE COUNT 226 a 136 c 165 g 177 t 1 others

ORIGIN

Query Match 92.2%; Score 320.8; DB 13; Length 705;
Best Local Similarity 99.1%; Pred. No. 9.8e-37;
Matches 322; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TGAACATCACTGCTTGGCAGAGTACCAACACTGTCATGTGATGATGCCGCCCTCT 60
|||||
DB 326 TGAACATCACTGCTTGGCAGAGTACCAACACTGTCATGTGATGATGCCGCCCTNT 267
|||||

QY 61 AGACCTACCCAGCGGACACATGCTCCGCTACCTTGGGTCTGAGGTCTGCTCAAG 120
|||||

DB 266 AGACCTACCCAGCGGACACATGCTCCGCTACCTTGGGTCTGAGGTCTGCTCAAG 207
|||||

QY 121 CGCTAGTGTAAAGCCGCTTGTGACAACTTAACCTACCTGCGCAAGAACAGTGTGGCC 180
|||||

DB 206 CGCTAGTGTAAAGCCGCTTGTGACAACTTAACCTACCTGCGCAAGAACAGTGTGGCC 147
|||||

QY 181 TTTTCGACCACTAGAACAACTTTTTCATTTGACAGTTGCGAATTTGTGAGTGTCTTTTA 240
|||||

DB 146 TTTTCGACCACTAGAACAACTTTTTCATTTGACAGTTGCGAATTTGTGAGTGTCTTTTA 87
|||||

QY 241 CATTTGATCTTTTGTGATGAGTTAGAGATGATTTTGCATGTATGCTTAATAAATCCT 300
|||||

DB 86 CATTTGATCTTTTGTGATGAGTTAGAGATGATTTTGCATGTATGCTTAATAAATCCT 27
|||||

QY 301 TGAATCATATAAAAAAAAAAAAAA 325

DB 26 TGAATCATATAAAAAAAAAAAAAA 2

RESULT 14
AA899757/c
LOCUS
DEFINITION
UI-R-E0-da-b-11-0-UI .s1 UI-R-E0 Rattus norvegicus cDNA clone
UI-R-E0-da-b-11-0-UI 3' similar to g114641851dbj1D25274/HUMPO2ST9
Homo sapiens mRNA, clone:PO2ST9, mRNA sequence.

AA899757
VERSION
AA899757.1 GI:4232251
KEYWORDS
SOURCE
ORGANISM
Rattus norvegicus (Norway rat)
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
COMMENT

1 (bases 1 to 465)
Bonaldo,M.F., Lennon,G. and Soares,M.B.
Normalization and subtraction: two approaches to facilitate gene discovery
Genome Res. 6 (9), 791-806 (1996)
97044477
8889548
On Apr 7, 1998 this sequence version replaced gi:3035111.
Contact: Soares, MB
Coordinated Laboratory for Computational Genomics
University of Iowa
375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: bento-soares@uiowa.edu
cDNA Library Preparation: M. Fatima Bonaldo, Ph.D. Clone
distribution: clones will be available through Research Genetics
Seq primer: M13 Forward.

FEATURES

source

1..465

location/Qualifiers

/organism="Rattus norvegicus"

/mol_type="mRNA"

/strain="Sprague-Dawley"

/db_xref="taxon:10116"

/clone="UI-R-E0-da-b-11-0-UI"

/dev_stage="embryonic"

/lab_host="DH10B (Life Technologies)"

/clone_lib="UI-R-E0"

/note="Vector: p773D-Pac (Pharmacia) with a modified polylinker: Site_1: NotI; Site_2: EcoRI. This library consists of a mixture of individually tagged normalized libraries constructed from 8, 12 and 18-day embryo. The tag is a string of 3-5 nucleotides present between the Not I site and the oligo-dT track which allows identification of the library of origin of a clone within the mixture."

BASE COUNT 155 a 93 c 99 g 118 t

ORIGIN

Query Match 91.4%; Score 318.2; DB 9; Length 465;
Best Local Similarity 99.1%; Pred. No. 2.9e-36;
Matches 320; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TGAACATCACTGCTTGGCAGAGTACCAACACTGTCATGTGATGATGCCGCCCTCT 60
|||||

DB 323 TGAACATCACTGCTTGGCAGAGTACCAACACTGTCATGTGATGATGCCGCCCTTT 264
|||||

QY 61 AGACCTACCCAGCGGACACATGCTCCGCTACCTTGGGTCTGAGGTCTGCTCAAG 120
|||||

DB 263 AGACCTACCCAGCGGACACATGCTCCGCTACCTTGGGTCTGAGGTCTGCTCAAG 204
|||||

QY 121 CGCTAGTGTAAAGCCGCTTGTGACAACTTAACCTACCTGCGCAAGAACAGTGTGGCC 180
|||||

DB 203 CGCTAGTGTAAAGCCGCTTGTGACAACTTAACCTACCTGCGCAAGAACAGTGTGGCC 144
|||||

OY 181 TTTCAGCACCTAGAACAACTTTTTCATTTGACAGTTCGAGATTTGGAGCTGTTT 240
|||||
DB 143 TTTCAGCACCTAGAACAACTTTTTCATTTGACAGTTCGAGATTTGGAGCTGTTT 84
OY 241 CATGATCTTTTGGTAAATGACAGTTCGAGTATTTGATGATGATTAATGCT 300
|||||
DB 83 CATGATCTTTTGGTAAATGACAGTTCGAGTATTTGATGATGATTAATGCT 24
OY 301 TGAATCATTAATAAAAAAAAAA 323
|||||
DB 23 TGAATCATTAATAAAAAAAAAA 1

RESULT 15
BE104696 405 bp mRNA linear EST 13-JUN-2000
LOCUS BE104696/C
DEFINITION UI-R-BX0-ars-e-01-0-UI-s1 UI-R-BX0 Rattus norvegicus cDNA clone
UI-R-BX0-ars-e-01-0-UI 3', mRNA sequence.
ACCESSION BE104696
VERSION BE104696.1 GI:8496796
KEYWORDS EST.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 405)
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene
discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
PUBMED 8889548

COMMENT
Contact: Soares, MB
Coordinated Laboratory for Computational Genomics
University of Iowa
375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: bento-soares@iowa.edu
The sequence contained an oligo-dT track that was present in the
oligonucleotide that was used to prime the synthesis of first
strand cDNA and therefore this may represent a bonafide poly A
tail. cDNA library Preparation: M.B. Soares lab Clone distribution:
clones will be available through Research Genetics (www.resgen.com)
The following repetitive elements were found in this cDNA sequence:
1-22, >AT-rich#Low-complexity
Seq primer: M13 Forward
POLY-A=yes.

FEATURES

source

1..405
Location/Qualifiers
/organism="Rattus norvegicus"
/mol_type="mRNA"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-BX0-ars-e-01-0-UI"
/dev_stage="embryonic 13 dpc"
/lab_host="DH10B (Life Technologies)"
/clone_lib="UI-R-BX0"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-BX0
library is derived from 13 dpc whole embryo tissue. For a
detailed description of the library from which this clone
was derived, please visit our web site at
ratseq.eng.uiowa.edu.
TAG_SEQ=None found"

BASE COUNT 131 a 76 c 87 g 110 t 1 others
ORIGIN

Query Match 91.1%; Score 317; DB 10; Length 405;
Best Local Similarity 98.5%; Pred. No. 4.6e-36;
Matches 320; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 1 TGAACATCATCTGCTCTTGGCAGAGTACCAACACTGTATGATGATGATGCCGCCCTCT 60
|||||
DB 328 TGAACATCATCTGCTCTTGGCAGAGTACCAACACTGTATGATGATGATGCCGCCCTCT 269
OY 61 AGACCTCAACCCACGGGGGACACATGCTCCGGTACCTTTGGGTCTGTAGAGTCTGTCAAG 120
|||||
DB 268 AGACCTCAACCCACGGGGGACACATGCTCCGGTACCTTTGGGTCTGTAGAGTCTGTCAAG 209
OY 121 CGCTAGTGTCTAACGGCGTTCTGTACAACTTACTGCTAGTGGCAAGACAGTGTGGGCC 180
|||||
DB 208 CGCTAGTGTCTAACGGCGTTCTGTACAACTTACTGCTAGTGGCAAGACAGTGTGGGCC 149
OY 181 TTTCAGCACCTAGAACAACTTTTTCATTTGACAGTTCGAGATTTGGAGCTGTTT 240
|||||
DB 148 TTTCAGCACCTAGAACAACTTTTTCATTTGACAGTTCGAGATTTGGAGCTGTTT 89
OY 241 CATGATCTTTTGGTAAATGACAGTTCGAGTATTTGATGATGATTAATGCT 300
|||||
DB 88 CATGATCTTTTGGTAAATGACAGTTCGAGTATTTGATGATGATTAATGCT 29
OY 301 TGAATCATTAATAAAAAAAAAA 325
|||||
DB 28 TGAATCATTAATAAAAAAAAAA 4

Search completed: August 25, 2003, 23:35:40
Job time : 727.909 secs

THIS PAGE BLANK (USPTO)


```

; TITLE OF INVENTION: USES THEREOF
; FILE REFERENCE: CL000862
; CURRENT APPLICATION NUMBER: US/09/734,675
; CURRENT FILING DATE: 2000-12-13
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 38844
; TYPE: DNA
; ORGANISM: Human
; US-09-734-675-3

Query Match          12.4%; Score 43; DB 4; Length 38844;
Best Local Similarity 55.8%; Pred. No. 0.043;
Matches 82; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

QY 201 TTTTTCATTTGACAGTTCGAGATTTGAGTGTTCATTTGATCTTTGCTAATGC 260
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 24453 TTCTTAAAAAAGCAATTTTACATTTAGGACTGTGTTAAAAAGCATTTTAAATATCA 24512

QY 261 AGTTACAGATGTTTTCATGATGACTTAATTAATCCTTGATCATATAAAAAAAA 320
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 24513 TTATTCAGAGAGTTTTCAGAGTTTTCATTTAAACCTTTAAACCAAAAAAAA 24572

QY 321 AAAAATGTCCTTGGAACTTGAAAAAA 347
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 24573 AAAGATTATGTGAAATTCGAACTAAA 24599

RESULT 3
US-07-867-106-2/c
; Sequence 2, Application US/07867106
; Patent No. 5389526
; GENERAL INFORMATION:
; APPLICANT: Slade, Martin B
; APPLICANT: Chang, Andy C M
; APPLICANT: Williams, Keith L
; TITLE OF INVENTION: Improved Plasmid Vectors for Cellular
; TITLE OF INVENTION: Slime Moulds of the genus Dictyostelium
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 5389526rls
; STREET: One Liberty Place 46th Floor
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; FILING DATE: US/07/867,106
; APPLICATION NUMBER: 19920625
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: AU PJ 7187
; APPLICATION NUMBER: PCT/AU90/00530
; FILING DATE: 02-NOV-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Feeney, Joanne Longo
; REGISTRATION NUMBER: 35,134
; REFERENCE/DOCKET NUMBER: RICE-0002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-568-3100
; TELEFAX: 215-568-3439
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5852 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)

```

```

; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 2378..5038
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 2378..5038
; US-07-867-106-2

Query Match          11.7%; Score 40.8; DB 1; Length 5852;
Best Local Similarity 54.7%; Pred. No. 0.094;
Matches 81; Conservative 0; Mismatches 67; Indels 0; Gaps 0;

QY 201 TTTTTCATTTGACAGTTCGAGATTTGAGTGTTCATTTGATCTTTGCTAATGC 260
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 5767 TTAAATTAATTTGTAATTTGTAATTTTATATATGTAATTTGTTGTTTACTT 5708

QY 261 AGTTACAGATGTTTTCATGATGACTTAATTAATCCTTGATCATATAAAAAAAA 320
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 5707 ATATTCTATTTTATTATTAATAATTAATTAATTAATTAATTAATTAATTAATTA 5648

QY 321 AAAAATGTCCTTGGAACTTGAAAAAA 348
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 5647 AAAAAAATTTAATTAATTAATTAATTA 5620

```

```

RESULT 4
US-08-345-756-8/c
; Sequence 8, Application US/08345756
; Patent No. 5633438
; GENERAL INFORMATION:
; APPLICANT: Baszczyński, Chris
; APPLICANT: Barbour, Eric
; APPLICANT: Hattori, Jiro
; APPLICANT: Mikl, Brian
; TITLE OF INVENTION: MICROSPORE-SPECIFIC REGULATORY ELEMENT
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/345,756
; FILING DATE: 22-NOV-1994
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, STEPHEN A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 33229/236/PIHI
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 859 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-345-756-8

Query Match          11.3%; Score 39.2; DB 1; Length 859;
Best Local Similarity 54.1%; Pred. No. 0.14;
Matches 80; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

```


OY	197	AAACCTTTTCAATGACAGATTGCAGAAATGTGGAGTGTTTTACATTATCATTTCGTA	256
Dd	240	AATTTTGGTTGAATTTTGTTCGAATCATGTGGGCTCCATTCAGTTGAGTTCGGGTA	181
OY	257	ATCGAGTAGCAGATGATTTTGCATGTATGACTTATATAATCCTGATCATAAAAAAA	316
Dd	180	CCCATTTTAATTAATGTATTTTCTTAAAAAATTCAAATATACCTTAATATGTCANATCA	121
OY	317	AAAAAAAAAATGCTTTGCAACTGAAAA	344
Dd	120	AAAAATATAATATATAAACCTATAAAA	93

RESULT 5
US-08-625-198-8/c

```

1  GENERAL INFORMATION:
2  APPLICANT: Baszczyński, Chris
3  APPLICANT: Barbour, Eric
4  APPLICANT: Hattori, Jiro
5  APPLICANT: Mikl, Brian
6  TITLE OF INVENTION: MICROSPORE-SPECIFIC REGULATORY ELEMENT
7  NUMBER OF SEQUENCES: 8
8  CORRESPONDENCE ADDRESS:
9  ADDRESSEE: Foley & Lardner
10 STREET: 3000 K Street, N.W., Suite 500
11 CITY: Washington
12 STATE: D.C.
13 COUNTRY: USA
14 ZIP: 20007-5109
15
16 COMPUTER READABLE FORM:
17 MEDIUM TYPE: Floppy disk
18 COMPUTER: IBM PC compatible
19 OPERATING SYSTEM: PC-DOS/MS-DOS
20 SOFTWARE: Patent Release #1.0, Version #1.30
21
22 CURRENT APPLICATION DATA:
23 APPLICATION NUMBER: US/08/625,198
24 FILING DATE: 01-APR-1996
25 CLASSIFICATION: 435
26
27 PRIOR APPLICATION DATA:
28 APPLICATION NUMBER: US 08/345,756
29 FILING DATE: 22-NOV-1994
30
31 ATTORNEY/AGENT INFORMATION:
32 NAME: BENT, STEPHEN A.
33 REGISTRATION NUMBER: 29,768
34 REFERENCE/DOCKET NUMBER: 33329/236/PIHI
35
36 TELECOMMUNICATION INFORMATION:
37 TELEPHONE: (202)672-5300
38 TELEFAX: (202)672-5399
39
40 TELEX: 904136
41
42 INFORMATION FOR SEQ ID NO: 8:
43 SEQUENCE CHARACTERISTICS:
44 LENGTH: 859 base pairs
45 TYPE: nucleic acid
46 STRANDEDNESS: single
47 TOPOLOGY: linear
48
49 MOLECULE TYPE: DNA (genomic)
50
51 US-08-625-198-8

```

Query Match	11.3%	Score 39.2	DB 1	length 85;
Best Local Similarity	54.1%	Pred. No. 0.14		
Matches 80; Conservative	0;	Mismatches 68;	Indels 0;	Gaps 0;

Oy 197 AACCTTTTTCATTCAGACGTGGCAGAAATGGGAGGTGTTTAACTATATCTTTGCA 256
 240 AATTTTGGTTCAATTTTGTTCGATCATTGGGGTTCATTCAGACTTCGAGTCCGGTA 181
 Oy 257 ATGCAGTTAGCAGTATGTTTTCAGATGATGACTTAATAATCCTTGAATCAATAAAAAA 316
 180 CCGATTATATATGATATTTTCTTAAAAAAATTCAAATATCACTTAATGCTCAAAATCA 121
 Db

```

QY      317 AAAAAAAAAATGCTTTGGAACTGAAA 344
          ||||| ||| | ||| | |||
Db      120 AAAAAATAAAATATATATAAACATAAAA 93

```

RESULT 6
US-08-345-756-5/c
; Sequence 5, Application US/08345756
; Patent No. 5633438
CURRENT INFORMATION

APPLICANT: Baszczynski, Chris
 APPLICANT: Barbour, Eric
 APPLICANT: Hattori, Jiro
 APPLICANT: Miki, Brian
 TITLE OF INVENTION: MICROSPORE-SPECIFIC REGULATORY ELEMENTS
 NUMBER OF SEQUENCES: 8
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Foley & Lardner
 STREET: 3000 K Street, N.W., Suite 500
 CITY: Washington
 STATE: D.C.
 COUNTRY: USA
 ZIP: 20007-5109
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/345,756
 FILING DATE: 22-NOV-1994
 CLASSIFICATION: 800
 ATTORNEY/AGENT INFORMATION:
 NAME: BENT, STEPHEN A.
 REGISTRATION NUMBER: 29,768
 REFERENCE/DOCKET NUMBER: 33329/236/PIH1
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (202)672-5300
 TELEFAX: (202)672-5399
 TELEX: 904136
 INFORMATION FOR SEQ ID NO: 5:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 2039 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: DNA (genomic)
 US-08-345-756-5

Query Match	11.3%	Score	39.2	DB	1	Length	2039
Best Local Similarity	54.1%	Pred. No.	0.18				
Matches	80	Conservative	0	Mismatches	68	Indels	0
						Gaps	0

OY	137	AAACCTTTTTCATTTGACAGTTCAGAGNAATGGAGATGTTTATACATTTTCGTA	256
Db	231	AATTTGGTTTGAATTTTCTTGCAATCATGTGGGTCATTCAGATTCAGATTCGGGTA	172
OY	257	ATGCATGTAGCAGTNGTTTGCATGATGACTTATTAATCCTGAATCATTAATAAAAA	316
Db	171	CCCATTTTATATGATATTTTCTTAAAAAATTCAAATATACTTAAATCGTCAAAATCA	112
OY	317	AAAAAAAATGCTTTGGAACTTGAATA	344
Db	111	AAAAATTAATAATATAAACAATAAAAA	84

RESULT 7
US-08-625-198-5/c
; Sequence 5, Application US/08625198
; Patent No. 5756324
; GENERAL INFORMATION:
; APPLICANT: Basczynski, Chris
; APPLICANT: Barbour, Eric


```

DB          2181 AAAAAAAAAAAAAAAAAAAAAA 2198

RESULT 10
US-09-277-716-17
: Sequence 17, Application US/09277716A
: Patent No. 6232107
: GENERAL INFORMATION:
: APPLICANT: Bryan, Bruce
: APPLICANT: Szent-Gyorgyi, Christopher
: APPLICANT: PROLUME, LTD.
: TITLE OF INVENTION: LUCIFERASES, FLUORESCENT PROTEINS, NUCLEIC ACIDS ENCODING THE
: CURRENT APPLICATION NUMBER: US/09/277,716A
: CURRENT FILING DATE: 1999-03-26
: EARLIER APPLICATION NUMBER: 60/102,939
: EARLIER FILING DATE: 1998-10-01
: EARLIER APPLICATION NUMBER: 60/089,367
: EARLIER FILING DATE: 1998-06-15
: EARLIER APPLICATION NUMBER: 60/079,624
: EARLIER FILING DATE: 1998-03-27
: NUMBER OF SEQ ID NOS: 32
: SOFTWARE: Patentln Ver. 2.0
: SEQ ID NO 17
: LENGTH: 1217
: TYPE: DNA
: ORGANISM: Renilla mulleri
: FEATURE:
: NAME/KEY: CDS
: LOCATION: (31)..(963)
: FEATURE:
: OTHER INFORMATION: Renilla mulleri luciferase
US-09-277-716-17

Query Match          10.6%; Score 36.8; DB 3; Length 1217;
Best Local Similarity 63.6%; Pred. No. 0.69;
Matches 56; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

Qy      238 TTACATGATCTTTTGCTAATGCAGTAGCAGTATGTTTGCAGTAGACTTAATAAT 297
      11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
Db      1136 TTCTCTTGCTTGCTATTATATAATACACATCTTCTATGTAATAAACTTGTTAATAAT 1185
      11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11

Qy      298 CCTGATCATATAAAAAAAAAAAAAA 325
      11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
Db      1186 TAATGATTCAGAAAAAAAAAAAAAAAAA 1213

RESULT 11
US-09-609-161B-17
: Sequence 17, Application US/09609161B
: Patent No. 6436682
: GENERAL INFORMATION:
: APPLICANT: Bryan, Bruce
: APPLICANT: Szent-Gyorgyi, Christopher
: APPLICANT: PROLUME, LTD.
: TITLE OF INVENTION: LUCIFERASES, FLUORESCENT PROTEINS, NUCLEIC ACIDS ENCODING THE LUC
: TITLE OF INVENTION: AND FLUORESCENT PROTEINS AND THE USE THEREOF IN DIAGNOSTICS, HIG
: TITLE OF INVENTION: SCREENING AND NOVELTY ITEMS
: FILE REFERENCE: 24729-121B
: CURRENT APPLICATION NUMBER: US/09/609,161B
: CURRENT FILING DATE: 2000-06-30
: PRIOR APPLICATION NUMBER: 09/277,716
: PRIOR FILING DATE: 1999-03-26
: PRIOR APPLICATION NUMBER: 60/102,939
: PRIOR FILING DATE: 1998-10-01
: PRIOR APPLICATION NUMBER: 60/089,367
: PRIOR FILING DATE: 1998-06-15
: PRIOR APPLICATION NUMBER: 60/079,624
: PRIOR FILING DATE: 1998-03-27
: NUMBER OF SEQ ID NOS: 32
: SOFTWARE: Patentln Ver. 2.0
: SEQ ID NO 17
: LENGTH: 1217
: TYPE: DNA

```

```

; ORGANISM: Renilla mulleri
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (31)..(963)
; OTHER INFORMATION: Renilla mulleri luciferase
US-09-609-161B-17

Query Match
Best Local Similarity 10.6%; Score 36.8; DB 4; Length 1217;
Matches 56; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

OY 238 TTACATGTCATCTTTGGTCAATAGCAGTAGCATGTTCGCATGTATAGCTAATAAT 297
DB 1126 TTTCTCTAGCTATTATTAATATCACACTCTGTATGTAAACTTGTTTATAAAT 1185
OY 298 CCTGAATCATATAAAAAAAAAAAAAA 325
DB 1186 TAATGATTCAAAAAAAAAAAAAAAAA 1213

RESULT 12
US-08-745-995A-31
; Sequence 31, Application US/08745995A
; Patent No. 6372887
; GENERAL INFORMATION:
; APPLICANT: Silver, Gary M.
; TITLE OF INVENTION: No. 6372887el Serine Protease Inhibitor
; TITLE OF INVENTION: Nucleic Acid Molecules, Proteins
; TITLE OF INVENTION: and Uses Thereof
; NUMBER OF SEQUENCES: 41
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Heeska Corporation
; STREET: 1825 Sharp Point Drive
; CITY: Fort Collins
; STATE: Colorado
; COUNTRY: USA
; ZIP: 80525
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: Windows 95
; SOFTWARE: WordPerfect for windows, Version 7.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/745, 995A
; FILING DATE: 07-NOV-1996
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Verser, Carol Talkington
; REGISTRATION NUMBER: 37,455
; REFERENCE/DOCKET NUMBER: FC-2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 970/493-7272
; TELEFAX: 970/484-9505
; INFORMATION FOR SEQ ID NO: 31:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1454 nucleotides
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 20..1210
US-08-745-995A-31

Query Match 10.5%; Score 36.4; DB 4; Length 1454;
Best Local Similarity 58.2%; Pred. No. 0.93;
Matches 64; Conservative 0; Mismatches 46; Indels 0; Gaps 0;

OY 216 GTTCAGAAATGTGAGATGTTTTACATGTATCTTTGCTAATCAGTAGCAGTATGTT 275
DB 1345 GTTACAGAAAATGTTTGGTTTGTAGTTTTCACCTTTTATGAATGAAATCATTATTAAT 1404

```


FILED DATE: 07-NOV-1996
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Verser, Carol Talkington
REGISTRATION NUMBER: 37,459
REFERENCE/DOCKET NUMBER: FC-2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 970/493-7272
TELEFAX: 970/484-9505
INFORMATION FOR SEQ ID NO: 33:
SEQUENCE CHARACTERISTICS:
LENGTH: 1454 nucleotides
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-09-005-352-33

Query Match 10.5%; Score 36.4; DB 4; Length 1454;
Best Local Similarity 58.2%; Pred. No. 0.93; Mismatches 0; Gaps 0;
Matches 64; Conservative 0; Indels 46; Gaps 0;
QY 216 GTTCAGAAATGTCAGATGTTTACATTCATCTTTGCTAATGACAGTTAGCAGTATGTT 275
DB 110 GTAGAGGAAAAAGTTGTTTGTAGTTTCACCTTTTATGATGATCACCCTATATAT 51
QY 276 TTGCATGTATGACTTAATAATCCCTGATCATTAATAAAAAAAAAAAAAAAAA 325
DB 50 GTGTAGTTTATGTAATAATAATGTAAATGTGAAAAAAAAAAAAAAAAAAAA 1

Search completed: August 26, 2003, 02:33:16
Job time: 26.5279 secs

Result No.	Score	Query Match	Length	DB	ID	Description
1	211.6	60.8	2941	14	US-10-198-846-9874	Sequence 9874, App
2	209.4	60.2	2051	9	US-09-923-302-255	Sequence 255, App
3	192.4	55.3	1644	14	US-10-153-668-48	Sequence 48, App
4	192.4	55.3	1646	14	US-10-153-668-46	Sequence 46, App
5	187.4	53.9	1232	10	US-09-880-107-1632	Sequence 1632, App
6	163.6	47.0	353	10	US-09-960-352-11432	Sequence 11432, App
7	116.4	33.4	3740	11	US-09-764-891-9984	Sequence 9984, App
8	95.4	27.4	201	10	US-09-960-332-4357	Sequence 4357, App
9	52.4	15.1	162	10	US-09-920-300A-1549	Sequence 1549, App
10	52.4	15.1	162	13	US-10-033-528-1549	Sequence 1549, App
11	52.4	15.1	477	14	US-10-066-543-3022	Sequence 3022, App
12	52.4	15.1	569	14	US-10-066-543-2878	Sequence 2878, App
13	52.4	15.1	547	14	US-10-066-543-2964	Sequence 2964, App
14	52.4	15.1	594	14	US-10-066-543-2864	Sequence 2864, App
15	52.4	15.1	605	11	US-09-871-161-483	Sequence 483, App
16	51.8	14.9	2313	14	US-10-198-846-12919	Sequence 12919, App

C 17	50.8	14.6	602	10	US-09-878-178-1654	Sequence 1654, App
C 18	50.8	14.6	602	13	US-10-046-935-1654	Sequence 1654, App
C 19	50.8	14.6	602	14	US-10-146-502-1654	Sequence 1654, App
C 20	43.2	12.4	424	10	US-09-960-352-11218	Sequence 11218, App
C 21	43	12.4	38844	13	US-10-060-333-3	Sequence 3, Appl1
C 22	42.2	12.1	640681	11	US-09-790-988-1	Sequence 1, Appl1
C 23	42	12.1	156361	11	US-09-764-891-10186	Sequence 10186, App
C 24	41.8	12.0	9817	12	US-10-311-455-1341	Sequence 1341, App
C 25	41.4	11.9	6464	12	US-10-311-455-487	Sequence 487, App
C 26	40.4	11.6	469	11	US-09-918-995-13017	Sequence 13017, App
C 27	39.4	11.3	6059	12	US-10-311-455-786	Sequence 786, App
C 28	39.2	11.3	277	10	US-09-960-352-12673	Sequence 12673, App
C 29	39.2	11.3	375	10	US-09-960-352-15014	Sequence 15014, App
C 30	39.2	11.3	6621	12	US-10-311-455-886	Sequence 886, App
C 31	39	11.2	501	13	US-10-073-632-184148	Sequence 184148, App
C 32	39	11.2	1460	14	US-10-198-846-9893	Sequence 9893, App
C 33	38.8	11.1	760	14	US-10-023-882-232	Sequence 232, App
C 34	38.8	11.1	2218	14	US-10-023-882-103	Sequence 103, App
C 35	38.8	11.1	2261	11	US-09-764-891-10185	Sequence 10185, App
C 36	38.6	11.1	1439	14	US-10-198-846-9746	Sequence 9746, App
C 37	38.2	11.0	5917	14	US-10-087-464-9	Sequence 9, Appl1
C 38	38	10.9	541	14	US-10-066-843-772	Sequence 772, App
C 39	38	10.9	871	14	US-10-198-846-8920	Sequence 8920, App
C 40	37.8	10.9	406	11	US-09-918-995-7460	Sequence 7460, App
C 41	37.8	10.9	408	10	US-09-960-352-6263	Sequence 6263, App
C 42	37.8	10.9	425	10	US-09-834-975-451	Sequence 451, App
C 43	37.8	10.9	5754	12	US-10-311-455-1494	Sequence 1494, App
C 44	37.8	10.9	1711	12	US-10-311-455-658	Sequence 628, App
C 45	37.6	10.8	2389	10	US-09-925-300-95	Sequence 95, Appl1

ALIGNMENTS

```

RESULT 1
US-10-198-846-9874
; Sequence 9874, Application US/10198846
; Publication No. US2003009974A1
GENERAL INFORMATION:
APPLICANT: Lillie, James
APPLICANT: Xu, Yongyao
APPLICANT: Wang, Youzhen
APPLICANT: Steinmann, Kathleen
TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS
TITLE OF INVENTION: FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
TITLE OF INVENTION: THERAPY OF BREAST CANCER
FILE REFERENCE: MRI-049
CURRENT APPLICATION NUMBER: US/10/198,846
CURRENT FILING DATE: 2002-07-18
PRIOR APPLICATION NUMBER: 60/306,220
PRIOR FILING DATE: 2001-07-18
NUMBER OF SEQ ID NOS: 14084
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 9874
LENGTH: 2941
TYPE: DNA
ORGANISM: Homo sapiens
US-10-198-846-9874

Query Match          60.8%; Score 211.6; DB 14; Length 2941;
Best Local Similarity 80.2%; Pred. No. 1.1e-45;
Matches 295; Conservative 0; Mismatches 49; Indels 24; Gaps 3

4 AACCTACTGCTTGTCGACAGATACCAACACTGTCATGATGTGATGGCGGCCCTCTAGA 63
|||||
Db 2106 AACATCTACTGCTTGTCGACAGATTAACCGACACTGTCACTTGACCAATTAAGA-CCTCTTTTA 2166
|||||

64 CCTCACCCACGGCGGACACATGCTCTCCGTA-----CCTTTGGGT 1020
|||||
Db 2165 CCTCGGCCACGGGAGACACAGCGCTCTGAGTCGCTTGCTTATGATGTTCTTTGGGT 2222
|||||

103 CTGTGAGGTTTGTCTCA--GGCTTATGCTAAAGCGGCTTCTGTGTCAACCTAATCACTACG 1600
|||||

```

Db 2225 CTGTGAGTCTGTAACTGTGCTAGTGTGCTGACAGTCTTCTGTACAACTTAACCTCACTGG 2284
QY 161 CAAGAACACAGTGTGGGCTTTGCGACCACTAGAACAACTTTTTCATTTGACAGTTGC 220
Db 2285 CGAGAAATACAGCCCTGGGACCTTCAGCCACTACAGAAATTTTAAATTTGACAGTTGC 2344
QY 221 AGAATTTGGAGTGTGTTTACATGATGATCTTGTGTAATGACAGTAATGTTTGA 280
Db 2345 AGAATTTGGAGTGTGTTTACATGATGATCTTGTGTAATGACAGTAATGTTTGA 2404
QY 281 TGTATGACTTAATTAATCCTTGAATCATTAATAAAAAAAAAAATGCTTTGGAACTTG 340
Db 2405 TGTATGACTTAATTAATCCTTGAATCATTAATAAAAAAAAAAATGCTTTGGAACTTG 2464
QY 341 AAAAAAAAA 348
Db 2465 AAAAAAAAA 2472

RESULT 2

US-09-925-302-255
; Sequence 255, Application US/09925302
; Patent No. US20020044941A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA104
; CURRENT APPLICATION NUMBER: US/09/925,302
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05918
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 896
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 255
; LENGTH: 2051
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (2)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (50)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (68)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (2027)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (2046)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-925-302-255

Query Match

Best Local Similarity 82.6%; Score 209.4; DB 9; Length 2051;
Matches 285; Conservative 0; Mismatches 36; Indels 24; Gaps 3;

QY 4 AACATCAGTGTCTGGCAGAGTACCAACACTGTCATGTGATGTCGCCGCCCTCTAGA 63
Db 1666 AACATCAGTGTCTGGCAGAGTACCAACACTGTCATGTGATGTCGCCGCCCTCTAGA 1724
QY 64 CACTCACCACGCGGACACATGCTTCCGCTA-----CCTTTGGGT 102
Db 1725 CACTCACCACGCGGACACAGCCCTCTGATGTCCTTGGCTATGATGTTCCCTTGGGT 1784
QY 103 CTGTGAGTCTGTGCAA--GCGCTAGTGTAAAGCCCTTCTGTAACTTAACCTCACTGG 160
Db 1785 CTGTGAGTCTGTGCAA--GCGCTAGTGTAAAGCCCTTCTGTAACTTAACCTCACTGG 1844

QY 161 CAAGAACACAGTGTGGGCTTTGCGACCACTAGAACAACTTTTTCATTTGACAGTTGC 220
Db 1845 CGAGAAATACAGCCCTGGGACCTTCAGCCACTACAGAAATTTTAAATTTGACAGTTGC 1904
QY 221 AGAATTTGGAGTGTGTTTACATGATGATCTTGTGTAATGACAGTAATGTTTGA 280
Db 1905 AGAATTTGGAGTGTGTTTACATGATGATCTTGTGTAATGACAGTAATGTTTGA 1964
QY 281 TGTATGACTTAATTAATCCTTGAATCATTAATAAAAAAAAAAATGCTTTGGAACTTG 325
Db 1965 TGTATGACTTAATTAATCCTTGAATCATTAATAAAAAAAAAAATGCTTTGGAACTTG 2009

RESULT 3

US-10-153-668-48
; Sequence 48, Application US/10153668
; Publication No. US20030092616A1
; GENERAL INFORMATION:
; APPLICANT: HONDA, Goichi
; APPLICANT: MATSUDA, Akio
; APPLICANT: MURAMATSU, Shuji
; APPLICANT: ISHIZAWA, Kenya
; TITLE OF INVENTION: STATE Activating Gene
; FILE REFERENCE: 1254-0207P
; CURRENT APPLICATION NUMBER: US/10/153,668
; PRIOR FILING DATE: 2002-05-24
; PRIOR APPLICATION NUMBER: US 60/293,172
; PRIOR FILING DATE: 2001-05-25
; PRIOR APPLICATION NUMBER: US 60/316,031
; PRIOR FILING DATE: 2001-08-31
; PRIOR APPLICATION NUMBER: US 60/328,403
; PRIOR FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: JP 2001-157043
; PRIOR FILING DATE: 2001-05-25
; PRIOR APPLICATION NUMBER: JP 2001-260681
; PRIOR FILING DATE: 2001-08-30
; PRIOR APPLICATION NUMBER: JP 2001-313175
; PRIOR FILING DATE: 2001-10-10
; NUMBER OF SEQ ID NOS: 488
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 48
; LENGTH: 1644
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (109)..(447)
US-10-153-668-48

Query Match

Best Local Similarity 81.7%; Score 192.4; DB 14; Length 1644;
Matches 268; Conservative 0; Mismatches 36; Indels 24; Gaps 3;

QY 4 AACATCAGTGTCTGGCAGAGTACCAACACTGTCATGTGATGTCGCCGCCCTCTAGA 63
Db 1318 AACATCAGTGTCTGGCAGAGTACCAACACTGTCATGTGATGTCGCCGCCCTCTAGA 1376
QY 64 CACTCACCACGCGGACACATGCTTCCGCTA-----CCTTTGGGT 102
Db 1377 CACTCACCACGCGGACACAGCCCTCTGATGTCCTTGGCTATGATGTTCCCTTGGGT 1436
QY 103 CTGTGAGTCTGTGCAA--GCGCTAGTGTAAAGCCCTTCTGTAACTTAACCTCACTGG 160
Db 1437 CTGTGAGTCTGTGCAA--GCGCTAGTGTAAAGCCCTTCTGTAACTTAACCTCACTGG 1496
QY 161 CAAGAACACAGTGTGGGCTTTGCGACCACTAGAACAACTTTTTCATTTGACAGTTGC 220
Db 1497 CGAGAAATACAGCCCTGGGACCTTCAGCCACTACAGAAATTTTAAATTTGACAGTTGC 1556
QY 221 AGAATTTGGAGTGTGTTTACATGATGATCTTGTGTAATGACAGTAATGTTTGA 280
Db 1557 AGAATTTGGAGTGTGTTTACATGATGATCTTGTGTAATGACAGTAATGTTTGA 1616

Query Match	15.1%;	Score 52.4;	DB 14;	Length 567;
Best Local Similarity	81.1%;	Pred. No. 0.00051;		

```
QY      4  AACATCACTGTCTTGGCAGAGTACCAACACTGTGATGTGATGCGCCCTCTAGA 63
      162 AACATCACTGTCTTGGCAGAGTACCGACACTGTGACCAATACTGA -CCCTCTTGA 104
Db      103 CCTGCCCGACGGGACACACGCGCTCTGTA 74

QY      64  CCTACCCAGCGGACACATGCTCCGGTA 93
      103 CCTGCCCGACGGGACACACGCGCTCTGTA 74

RESULT 14
US-10-066-543-2864
; Sequence 2864, Application US/10066543
; Publication No. US20030087818A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Yugu
; APPLICANT: Pyle, Ruth A.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Indrias, Carol Yoseph
; APPLICANT: Lodes, Michael J.
; APPLICANT: Secrist, Heather
; APPLICANT: Carter, Derrick
; APPLICANT: Fanger, Gary R.
; APPLICANT: Smith, Carole L.
; APPLICANT: Durham, Margareta
; APPLICANT: Stolk, John A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.563
; CURRENT APPLICATION NUMBER: US/10/066,543
; CURRENT FILING DATE: 2002-01-31
; NUMBER OF SEQ ID NOS: 3417
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2864
; LENGTH: 594
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-066-543-2864

Query Match      15.1%; Score 52.4; DB 14; Length 594;
Best Local Similarity 81.1%; Pred. No. 0.00052;
Matches 73; Conservative 0; Mismatches 16; Indels 1; Gaps 1;

QY      4  AACATCACTGTCTTGGCAGAGTACCAACACTGTGATGTGATGCGCCCTCTAGA 63
      436 AACATCACTGTCTTGGCAGAGTACCGACACTGTGACCAATACTGA -CCCTCTTGA 494
Db      495 CCTGCCCGACGGGACACACGCGCTCTGTA 524

QY      64  CCTACCCAGCGGACACATGCTCCGGTA 93
      103 CCTGCCCGACGGGACACACGCGCTCTGTA 74
Db      495 CCTGCCCGACGGGACACACGCGCTCTGTA 524

RESULT 15
US-09-871-161-483/C
; Sequence 483, Application US/09871161
; Publication No. US20030097666A1
; GENERAL INFORMATION:
; APPLICANT: ENDEGE, WILSON O., ET AL.
; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
; FILE REFERENCE: CCDNA-260XX
; CURRENT APPLICATION NUMBER: US/09/871,161
; CURRENT FILING DATE: 2001-05-31
; PRIOR APPLICATION NUMBER: 09/328,111
; PRIOR FILING DATE: 1999-06-08
; PRIOR APPLICATION NUMBER: 60/117,393
; PRIOR FILING DATE: 1999-01-27
; PRIOR APPLICATION NUMBER: 60/098,639
; PRIOR FILING DATE: 1998-08-31
; NUMBER OF SEQ ID NOS: 544
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 483
; LENGTH: 605
```

```
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1)-(605)
; OTHER INFORMATION: n = A,T,C or G
US-09-871-161-483

Query Match      15.1%; Score 52.4; DB 11; Length 605;
Best Local Similarity 81.1%; Pred. No. 0.00052;
Matches 73; Conservative 0; Mismatches 16; Indels 1; Gaps 1;

QY      4  AACATCACTGTCTTGGCAGAGTACCAACACTGTGATGTGATGCGCCCTCTAGA 63
      162 AACATCACTGTCTTGGCAGAGTACCGACACTGTGACCAATACTGA -CCCTCTTGA 104
Db      103 CCTGCCCGACGGGACACACGCGCTCTGTA 74

QY      64  CCTACCCAGCGGACACATGCTCCGGTA 93
      103 CCTGCCCGACGGGACACACGCGCTCTGTA 74
Db      103 CCTGCCCGACGGGACACACGCGCTCTGTA 74

Search completed: August 26, 2003, 02:45:57
Job time : 86.049 secs
```